VOL. 46 1955

The Journal of Educational Psychology

Devoted Primarily to the Scientific Study of Problems of Learning and Teaching

4782

BOARD OF EDITORS*

WM. CLARK TROW, *Chairman*
University of Michigan

STEPHEN M. COREY
Teachers College
Columbia University

J. B. STROUD
State University of Iowa

---

*Appointed by the American Psychological Association

EDITORIAL CONSULTANTS

PAUL BLOMMERS
CHESTER W. HARRIS
HAROLD E. JONES

H. H. REMMERS
PERCIVAL M. SYMONDS

MILES A. TINKER
ALEXANDER G. WESMAN
PAUL A. WITTY

Baltimore, Md.
WARWICK and YORK, Inc.
1955

11.8.70
9 103

TABLE OF CONTENTS

Abnormal Psychology as a Selective Factor in the College Curriculum. EUGENE S. MILLS 101–111
Adolescents' Views on One Aspect of their Development. RUTH STRANG 423–432
Application of Fisher's Discriminant Function in the Classification of Students, An. J. STANLEY AHMANN 184–188
Bilingualism and Biculturalism. JAMES P. SOFFIETTI .. 222–227
Book Reviews. 56–64, 121–128, 251–256, 311–320, 381–384, 446–448.
Characteristics of Poor and Good Readers as Disclosed by the Wechsler Intelligence Scale for Children, The. HAROLD F. BURKS AND PAUL BRUCE 488–493
Comparative Study of Values and Critical Thinking Skills of a Group of Educational Workers, A. JOSEPH C. BLEDSOE 408–417
Comparison of Davis-Eells and Kuhlmann-Finch Scores of Children from High and Low Socio-Economic Status, A. WILLIAM COLEMAN AND ANNIE W. WARD 465–469
Configurational Learning in Children. ALLEN D. CALVIN 117–120
Davis-Eells Games as a Measure of the Intelligence of Poor Readers, The. JOSEPH JUSTMAN AND MIRIAM ARONOW 418–422
Democratic College Student, The. MAURICE F. FREEHILL 477–487
Development of Shorter and More Useful Selection Tests, The. A. PEMBERTON JOHNSON 402–407
Dominant Attitudes of Negro Teachers toward Integration in Education, The. ROBERT T. AMOS 470–476
Drawing Technique for Studying Certain Outcomes of Teacher Education, A. WILLIAM RABINOWITZ AND ROBERT M. W. TRAVERS 257–273
Effect of Organizational Aids on Comprehension and Retention, The. C. M. CHRISTENSEN AND K. E. STORDAHL 65–74

Effect of Reading Ability on Two Standardized Classification Tests, The. CECIL J. MULLINS 189–192

Effects of Variations in Teacher Rôle Behavior on the Student-Teacher Relationship. MARTIN M. GROSSACK 433–436

Factor Analysis of the SRA Youth Inventory, A. JOHN P. PAISIOS AND H. H. REMMERS 25–30

Forced Choice Rating Form for College Teachers, A. ROLLIN M. PATTON AND PRISCILLA R. MEYER 499–503

Groupness of the Group. IRVING LORGE 449–456

Leadership Behavior Associated with the Administrative Reputation of College Departments. JOHN K. HEMPHILL 385–401

Mental Maturity versus Perception Abilities in Primary Reading. SISTER MARY JAMES HARRINGTON AND DONALD D. DURRELL 375–380

Normative Test Data for Adult Males Obtained by House-to-House Testing. ELIZABETH P. HAGEN AND ROBERT L. THORNDIKE 207–216

Note on the Standard Error of Measurement, A. C. H. PATTERSON 239–242

Note on the Validity of the Army General Classification Test as a Predictor of Academic Achievement. TOLAN L. CHAPPELL 53–55

Organismic Age Concept, The. PAUL BLOMMERS, LOTUS M. KNIEF, AND J. B. STROUD 142–150

Organismic Growth: Sexual Maturity and Progress in Reading. FRED T. TYLER 85–93

Personality and Teacher Effectiveness: A Problem in Theoretical Research. E. G. GUBA AND J. W. GETZELS 330–344

Predicting Attrition-Survival in First Year Engineering. JOHN P. MALLOY, BOLESLAW WYSOCKI, AND LEO F. GRAHAM 217–221

Prediction of the Academic Success of College Freshmen. ROBERT A. JACKSON 296–301

Preliminary Exploration of Personality Factors Related to Success in a High-Level Military School. JAMES H. L. ROACH AND EDWARD M. WALL 247–250

Problem-Solving by Teams and by Individuals in a Field

Setting. Irving Lorge, Jacob Tuckman, Louis Aikman, Joseph Spiegel, and Gilda Moss...... 160–166
Punitiveness and "Causalty" in Grade School Children. Eugene E. Levitt.......................... 494–498
Pupils' Values and the Validity of the Minnesota Teacher Attitude Inventory. G. M. Della Piana and N. L. Gage................................. 167–178
Reading Improvement and College Grades: A Follow-Up. Donald E. P. Smith and Roger L. Wood........ 151–159
Relationship of Immediate Retention of Technical Training Material to Career Preferences and Aptitudes, The. G. R. Klare, L. M. Gustafson, J. E. Mabry, and E. H. Shuford................. 321–329
Relationship of Style Difficulty to Immediate Retention and to Acceptability of Technical Material, The. George R. Klare, James E. Mabry, and Levarl M. Gustafson............................. 287–295
Relative Stability of Classroom Social Status with Unweighted and Weighted Sociometric Choices, The. Norman E. Gronlund...................... 345–354
Relativistic Definition of Stimulus-Response, A. Morris L. Bigge............................. 457–464
Retroactive Inhibition in Meaningful Material. John F. Hall................................. 47–52
Scoring Difficulty of Wechsler Comprehension Responses. Galen R. Plumb and Don C. Charles........ 179–183
Second Study of Characteristics of Good and Poor Spellers, A. David H. Russell............... 129–141
Self-Appraisal in Teacher Personality. Sister Mary Amatora.............................. 94–100
Sex Differences in Adjustment to College. John Summerskill and C. Douglas Darling............ 355-361
SHORTHAND STUDY AND SPELLING ABILITIES. Barry T. Jensen and Junne W. Jensen................ 112–116
Social Structure of the Classroom Group: An Observational Framework. Gale E. Jensen........... 362–374
Solutions by Teams and by Individuals to a Field Problem at Different Levels of Reality. Irving Lorge, Jacob Tuckman, Louis Aikman, Joseph Spiegel, and Gilda Moss........................ 17–24

Stability of Interest Scores of College Freshmen, The. CLAY E. GEORGE AND ALBERT J. KINGSTON 243–246
Student, Supervisor and Self-Ratings of Instructional Proficiency. WILSE B. WEBB AND C. Y. NOLEN .. 42–46
Study-Attitudes Questionnaire for Predicting Academic Success, A. WILLIAM F. BROWN AND WAYNE H. HOLTZMAN 75–84
Study of the Effects of Teacher Knowledge of Pupil Characteristics on Pupil Achievement and Attitudes Towards Classwork. KENNETH B. HOYT .. 302–310
Teacher Reactions to Conflict Situations. SAMUEL I. SPECTOR 437–445
Teamwork in the College Class. HENRY CLAY SMITH .. 274–286
Technique in Problem-Solving as a Predictor of Educational Achievement. K. PATRICIA CROSS AND EUGENE L. GAIER 193–206
Use of an Electronic Computer in Principal Axes Factor Analysis, The. CHARLES WRIGLEY AND JACK O. NEUHAUS 31–41
Use of Dual Grouping in Student-Centered Teaching, The. NORMAN E. GRONLUND 1–16
War-Time Accelerates Ten Years After. MARIE A. FLESHER AND S. L. PRESSEY 228–238

BOOK REVIEWS

Bruno Bettelheim. *Symbolic Wounds, Puberty Rites and the Envious Male.* (H. MELTZER) 254
Glenn Myers Blair, R. Stewart Jones, and Ray H. Simpson. *Educational Psychology.* (PERCIVAL M. SYMONDS AND ARTHUR R. JENSEN) 56
Sir Cyril Burt. *The Causes and Treatment of Backwardness.* (H. MELTZER) 252
Stephen M. Corey. *Action Research to Improve School Practices.* (MAURICE D. AND JEANNE A. WOOLF) 251
W. D. Commins and Barry Fagin. *Principles of Educational Psychology.* (PERCIVAL M. SYMONDS AND ARTHUR R. JENSEN) 53
L. J. CRONBACH. *Educational Psychology.* (PERCIVAL M. SYMONDS AND ARTHUR R. JENSEN) 56
Lee J. Cronbach. *Educational Psychology.* (ARTHUR P. COLADARCI) 383
John Dollard, Frank Auld, Jr., and Alice Marsden White. *Steps in Psychotherapy.* (PERCIVAL M. SYMONDS) 311
Marvin D. Glock. *The Improvement of College Reading.* (MILES A. TINKER) 255
J. Anthony Humphreys and Arthur E. Traxler. *Guidance Services.* (MAURICE D. AND JEANNE A. WOOLF) 127
Harrington V. Ingham and Leonore R. Love. *The Process of Psychotherapy.* (PERCIVAL M. SYMONDS) 311
Kentucky Symposium, The. *Learning Theory, Personality Theory and Clinical Research.* (A. S. EDWARDS) 446
George R. Klare and Byron Buck. *Know Your Reader.* (MILES A. TINKER) 318
Bruno Klopfer, Mary D. Ainsworth, Walter G. Klopfer, and Robert R. Holt. *Developments in Rorschach Technique. Volume I: Technique and Theory.* (LEE J. CRONBACH) 121
Henry Clay Lindgren. *Mental Health in Education.* (A. S. EDWARDS) 382
O. Hobart Mowrer, Editor. *Psychotherapy: Theory and Research.* (PERCIVAL M. SYMONDS) 311
Margaret Naumberg. *Psychoneurotic Art: Its Function in Psychotherapy.* (H. MELTZER) 381

Doris M. Odlum. *Psychology, The Nurse and the Patient.* (H. MELTZER) 446

H. H. Remmers, Einar R. Ryden, and Clellen L. Morgan. *Introduction to Educational Psychology.* (PERCIVAL M. SYMONDS AND ARTHUR R. JENSEN) 56

C. C. Ross. *Measurement in Today's Schools.* Revised by Julian C. Stanley, third edition. (LEE J. CRONBACH) .. 319

Henry P. Smith. *Psychology and Teaching.* (PERCIVAL M. SYMONDS, AND ARTHUR R. JENSEN) 56

Herbert Sorenson. *Psychology and Education,* 3rd Ed. (PERCIVAL M. SYMONDS AND ARTHUR R. JENSEN) 56

W. S. Taylor. *Dynamic and Abnormal Psychology.* (A. S. EDWARDS) 125

Charles B. Thompson and Alfreda P. Sill. *Our Common Neurosis.* (H. MELTZER) 123

Carl A. Whitaker and Thomas P. Malone. *The Roots of Psychotherapy.* (PERCIVAL M. SYMONDS) 311

Lewis R. Wolberg. *The Technique of Psychotherapy.* (PERCIVAL M. SYMONDS) 311

INDEX TO AUTHORS

Ahmann, J. Stanley 184
Aikman, Louis 17, 160
Amatora, Sister Mary 94
Amos, Robert T. 477
Aronow, Miriam 418
Bigge, Morris L. 457
Bledsoe, Joseph C. 408
Blommers, Paul 142
Brown, William F. 75
Bruce, Paul 488
Burks, Harold F. 488
Calvin, Allen D. 117
Chappell, Tolan L. 53
Charles, Don C. 179
Christensen, C. M. 65
Coleman, William 465
Cross, K. Patricia 193
Darling, C. Douglas 355
Della Piana, G. M. 167
Durrell, Donald D. 375
Flesher, Marie A. 228
Freehill, Maurice F. 477
Gage, N. L. 167
Gaier, Eugene L. 193
George, Clay E. 243
Getzels, J. W. 330
Graham, Leo F. 217
Gronlund, Norman E. 1, 345
Grossack, Martin M. 433
Guba, E. G. 330
Gustafson, Levarl M. 287, 321
Hagen, Elizabeth P. 207
Hall, John F. 47
Harrington, Sister Mary James 375
Hemphill, John K. 385
Holtzman, Wayne H. 75
Hoyt, Kenneth B. 302
Jackson, Robert A. 296
Jensen, Barry T. 112
Jensen, Gale E. 362
Jensen, Junne W. 112
Johnson, A. Pemberton 402
Justman, Joseph 418
Kingston, Albert J. 243
Klare, George R. 287, 321
Knief, Lotus M. 142
Levitt, Eugene E. 494
Lorge, Irving 17, 160, 449
Mabry, James E. 287, 321
Malloy, John P. 217
Meyer, Priscilla R. 499
Mills, Eugene S. 101
Moss, Gilda 17, 160
Mullins, Cecil J. 189
Neuhaus, Jack O. 31
Nolan, C. Y. 42
Paisios, John P. 25
Patterson, C. H. 239
Patton, Rollin M. 499
Plumb, Galen R. 179
Pressey, S. L. 228
Rabinowitz, William 257
Remmers, H. H. 25
Roach, James H. L. 247
Russell, David H. 129
Shuford, E. H. 321
Smith, Donald E. P. 151
Smith, Henry Clay 274
Soffietti, James P. 222
Spector, Samuel I. 437
Spiegel, Joseph 17, 160
Stordahl, K. E. 65
Strang, Ruth 423
Stroud, J. B. 142
Summerskill, John 355
Thorndike, Robert L. 207
Travers, Robert M. W. 257
Tuckman, Jacob 17, 160
Tyler, Fred T. 85
Wall, Edward M. 247
Ward, Annie W. 465
Webb, Wilse B. 42
Wood, Roger L. 151
Wrigley, Charles 31
Wysocki, Boleslaw 217

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume 46 January, 1955 Number 1

418 261

THE USE OF DUAL GROUPING IN STUDENT-CENTERED TEACHING

NORMAN E. GRONLUND

College of Education, University of Illinois

One of the persistent problems in student-centered teaching is that of shifting the major leadership responsibility from the teacher to the students as rapidly and efficiently as possible. Due to the lack of experience in democratic procedures the typical student is very dependent upon the teacher for detailed instructions concerning what to study, when to study and how to study. Consequently, some systematic procedure is necessary in assisting the student to move from this extreme state of dependency to that of an active group member willing to accept the major responsibility for planning and carrying out his learning experiences. This paper will describe how dual grouping was used in an Educational Psychology class as a method of providing such systematic organization to student-centered teaching.

The class consisted of twenty-seven students in an undergraduate Educational Psychology course. A modified problem[1] approach was used to study the application of psychological principles to teaching-learning situations. In general, the procedure consisted of identifying, selecting, analyzing, and developing possible solutions to teaching-learning problems that the students were most likely to face as professional teachers. The following list of objectives, which was given to each student at the beginning of the semester, will clarify the nature of the course in terms of specific outcomes that were expected.

---

[1] For a detailed description concerning the assumptions underlying this course approach as well as the procedures used, consult Ray H. Simpson, *Improving Teaching-Learning Processes*. New York: Longmans, Green & Co., 1953, pp. ix, 487.

COURSE OBJECTIVES

1) The ability to solve professional problems requiring a knowledge of psychological principles.

1.1. Skill in identifying professional problems related to the psychological aspects of teaching-learning situations.
  1.11. Recognizes problems that arise in teaching-learning situations.
  1.12. States problems clearly and objectively.
  1.13. Analyzes problems into workable sub-problems.
1.2. Skill in selecting professional problems upon which it is most desirable to work at a specified time.
  1.21. Constructs criteria for the selection of problems.
  1.22. Selects problems that meet criteria.
1.3. Skill in locating and selecting resources for the solution of professional problems.
  1.31. Locates a variety of resource material through the use of the *Educational Index*, *Review of Educational Research*, *Psychological Abstracts*, and other guides.
  1.32. Selects resource material which is pertinent to the problem being studied.
1.4. Skill in purposeful and functional professional reading.
  1.41. Reads selectively in terms of the pertinence of the material to the problem. Ignores unessentials.
  1.42. Interprets tables, charts and graphs that are pertinent to the problem.
  1.43. Evaluates the results of research in terms of the adequacy of the sample, the control of variables, and the validity of the conclusions.
1.5. Skill in maintaining a professional learning record that contributes to the solution of professional problems requiring a knowledge of psychological principles.
  1.51. Records material which is pertinent to the problem.
  1.52. Records material which reflects a variety of opinions and points of view.
  1.53. Gives analytical and critical reading reactions to recorded materials.
  1.54. Records material which is sufficient quantitatively to draw valid conclusions.
  1.55. Analyzes and integrates recorded material in summaries.
  1.56. States conclusions which are valid in terms of recorded material.
1.6. Skill in democratic, interpersonal relationships.
  1.61. Actively participates in group discussions.
  1.62. Makes constructive contributions to group work.
  1.63. Shares ideas and responsibilities with other group members.
1.7. Skill in self-evaluation, including self-diagnosis.
  1.71. Recognizes own strengths and limitations in ability to solve professional problems.

1.72. Recognizes own strengths and limitations in interpersonal relationships.
1.73. Objectively evaluates own performance in group work.

It will be noted that the previously mentioned objectives stress process skills which were common goals of all learners. Content objectives were intentionally omitted, since they were to be developed coöperatively by the learners and the instructor with the needs of each individual learner in mind. As developed later, some of these objectives, and their related activities, were those of individual learners, some of small groups, and some were common to the whole class. To add cohesiveness and unity to such diversified activities, as well as to shift from a teacher-centered class to a student-centered class, it seemed desirable to utilize some systematic grouping procedure. A dual grouping system, consisting of permanent work groups and committees, was finally devised.

FORMING PERMANENT WORK GROUPS AND COMMITTEES

During the first several weeks various types of groups were utilized. General class discussions were used in dealing with problems of common interest. When activities could be most efficiently carried out by small group work the class was divided up into groups ranging from four to six in number. At the beginning of the semester students were arbitrarily assigned to work groups. During the identification and selection of professional problems requiring a knowledge of psychological principles the students first met in groups according to their major teaching fields and later met in groups according to the professional problems they had selected for study. This flexibility in grouping, in addition to meeting the needs of the current activities in class, enabled the students to broaden their acquaintanceships among the class members.

When it was felt that students had had sufficient contact with other members of the class to make valid choices of work companions, permanent work groups and committees were formed. They were formed in the following manner:

1) *Students were asked to write the names of five class members with whom they most preferred to work for the remainder of the semester.*—It was emphasized that it would be impossible to arrange groups so that students would get all of their choices. However, each student was assured that his choices would be given first preference, mutual

choices would be considered next, and finally other students choosing him would be taken into account.

2) *The students' choices of work companions were analyzed by the instructor through the use of a matrix table.*[2] This analysis resulted in six permanent work groups. Four of the groups contained five students, one contained four students and one contained three students. An attempt was made to keep the work groups at five, since there were to be five committees, with each member of the work group serving on a committee. The value of this arrangement will be clarified when the functions of the work groups and the committees are discussed later. Since there were twenty-seven students in class, it appeared best to have two smaller groups of four and three students rather than only one small group of two students.

With the above arrangement of groups, sixty-four of the students' choices of work companions were satisfied. Consequently, the majority of students were placed in groups with two or more of the classmates they had chosen. In four instances it was necessary to place students in groups where they had chosen only one member of the group. However, in all four cases the students had received one or more choices from some member of the group to which they were assigned.

There were forty-eight reciprocated choices among the sixty-four student choices that were satisfied. In other words, the groups were extremely cohesive in social structure with a majority of mutual companions present in each group.

3) *Students met in the six permanent work groups, resulting from the sociometric analysis and appointed each member to one of the following committees: advisory, evaluation, presentation, resource, and visual aids.*—The nature and purpose of the dual grouping system was explained before committee appointments were made. It was pointed out that each member of a permanent work group would serve on a different committee. Consequently, each committee would contain one representative from each work group. This was desirable since the committee work would be concerned mainly with service activities related to the smooth functioning of classroom experiences and total representation was desired. To provide some basis for appointing committee members, the general re-

---

[2] Merle M. Ohlsen, "Helping teachers interpret sociometric test data," *The Journal of Teacher Education*, 2, June, 1951, pp. 99–104.

sponsibilities of each committee were outlined. It was understood that each committee would receive a tentative list of suggested responsibilities, as illustrated on the following pages, to provide further orientation when committee work began. At this time, the fact was stressed that the member of each work group appointed to the advisory committee would also serve as chairman of his work group. This was emphasized as a key appointment, since these students would be responsible for the leadership of their work groups, as well as representing them in planning future course activities with the instructor.

After the nature and purpose of dual grouping had been clarified the work groups were free to appoint each of their members to one of the committees. Some groups did this by voting, others through discussion and mutual agreement.

4) *Students met in committees and elected a chairman of each committee.*—As soon as the committees were appointed a meeting of each committee was held. During this first meeting the chairmen were elected and each committee began work on an assigned task. The instructor moved from committee to committee clarifying the nature of their first responsibility and helping them get started with it. The advisory committee assisted the instructor in planning the next class meeting. The evaluation committee began to set up criteria for evaluating student performance in permanent work groups. The presentation committee began to set up criteria for evaluating students' presentations in class. The resource committee began to compile a list of reviewing, abstracting and index services available in the library. The visual aids committee began reviewing the films that might be suitable for use in class. Committee work went ahead rapidly during this first meeting.

When permanent work groups and committees were formed, as described above, each member in class received a list of suggested responsibilities for these various groups, as well as a two-fold chart indicating the permanent work group and committee to which each student belonged. It was agreed that this tentative list of responsibilities would be reviewed and additions or modifications would be considered at future meetings.

SUGGESTED RESPONSIBILITIES OF PERMANENT WORK GROUPS AND COMMITTEES

The suggested responsibilities of the permanent work groups and the committees are listed below. Although they are stated in terms of specific group responsibilities, it is, of course, understood that these responsibilities are shared by the instructor through individual and group contact. In addition, many of these activities will be participated in by the entire class where such participation appears warranted. Work groups and committees should review their suggested responsibilities to determine what additions or modifications appear desirable.

Suggested responsibilities of permanent work groups.—The permanent work groups will continue with the functions previously handled by the informal classroom groups. These include the following responsibilities related to the identification, selection and solution of professional problems requiring a knowledge of psychological principles.

1) Mutual assistance in the further identification of professional problems requiring a knowledge of psychological principles.

2) Mutual assistance in the selection of professional problems upon which it is most desirable to work at a specified time.

3) Mutual assistance in breaking down professional problems selected into workable sub-problems.

4) Review criteria set up for the evaluation of professional problems and sub-problems.

5) Mutual assistance in the evaluation of each other's professional problems and sub-problems in terms of the above criteria.

6) Exchange of information concerning appropriate resources available for the development of solutions to professional problems.

7) Exchange of information on the development and maintenance of the professional learning record.

8) Set up cirteria for evaluation of the professional learning record.

9) Mutual assistance with a periodic evaluation of each other's professional learning record in terms of the above criteria.

10) Mutual assistance in the improvement of interpersonal relationships.

11) Mutual evaluation of the nature and extent of each member's contributions and the effectiveness of their interpersonal relationships in small group work.

12) Other responsibilities. (To be added by students.)

Suggested responsibilities of advisory committee.—The advisory committee is to assist the instructor in planning in-class and out-of-class activities. The following activities are suggestive of specific responsibilities in this regard:

1) Determine the specific purposes of each class meeting.

2) Decide on the type of assignment necessary to make each class meeting most meaningful.

3) Determine the type of class activities for each class meeting and the amount of time to be devoted to each activity. Suggestive class activities are class discussions, student presentations, small group work, individual work, and committee meetings.

4) Devise methods for evaluating the success of each class meeting. (This responsibility may be referred to the evaluation committee for suggestions and assistance.)

5) Set up a tentative schedule of class meetings for several weeks in advance.

6) Revise tentative class schedule in accordance with class members suggestions for the improvement of class meetings, and in accordance with the needs of students arising from the on-going teaching-learning process.

7) Other responsibilities. (To be added by students.)

Suggested responsibilities of evaluation committee.—The evaluation committee is responsible for setting-up criteria and devising instruments for the evaluation of student and instructor performance. The following activities are suggestive of specific responsibilities in this area:

1) Set up criteria for evaluating the performance of students in small group work (permanent work groups and committees).

2) Devise student rating scale in harmony with the above criteria.

3) Set up criteria for evaluating the performance of the instructor in all phases of in-class activity and related out-of-class activities, such as individual conferences, out-of-class group meetings, preparation, etc.

4) Devise teacher rating scale in harmony with the above criteria.

5) Provide other committees with assistance in setting up criteria or devising instruments of evaluation.

6) Administer student and teacher rating scales.

7) Other responsibilities. (To be added by students.)

Suggested responsibilities of presentation committee.—The presentation committee is the clearing house for all student presentations. Each student will have his presentation approved by this committee and then schedule a time for his presentation with the advisory committee. The following activities are suggestive of specific responsibilities of the presentation committee:

1) Set up criteria regarding the nature and content of an acceptable student presentation.

2) Set up an outline which students can use in applying for permission to make a class presentation. This would include such things as purpose of presentation, main points to be covered, method of presentation, etc.

3) Devise rating scale for evaluating the success of students' presentations.

4) Administer rating scale after each student's presentation.

5) Other responsibilities. (To be added by students.)

Suggested responsibilities of resource committee.—The resource committee is responsible for keeping the class members informed concerning the location, selection and utilization of resources related to the solution of professional problems requiring a knowledge of psychological principles. The suggested responsibilities of this committee are indicated below:

1) Survey the reviewing, abstracting, and index services available in the library, noting their location.

2) Compile a list of the above services available indicating the types of material that can be located with each, the frequency with which issues appear, and where they may be found in the library.

3) Prepare a brief list of suggestions for the most efficient use of the library.

4) Prepare a list of resources other than professional books and journals that might be utilized in the solution of professional problems.

5) Other responsibilities. (To be added by students.)

Suggested responsibilities of visual aids committee.—The visual aids committee is responsible for keeping the class members informed concerning the films available in educational psychology, the method of ordering the previewing films, and the principles underlying their effective use in the classroom. All students will clear with this committee if they use visual aids in their presentations. The suggested responsibilities of this committee are as follows:

1) Compile a list of available films in educational psychology.

2) Prepare a list of purposes for which visual aids might be used.

3) Prepare an outline of the necessary steps in ordering and previewing films from the Visual Aids Service.

4) Prepare a list of suggestions for the most efficient use of visual aids in the classroom.

5) Other responsibilities. (To be added by students.)

All of the above responsibilities are to be carried out under the guidance of the instructor. Criteria and instruments of evaluation constructed by committees are subject to revision and final approval by a majority of the class membership, as well as by the instructor.

This list of responsibilities may indicate too much structure for a student-centered class. However, it was stressed that these were merely suggested responsibilities to give work groups and committees something tangible to work on when first organized. This structuring did seem to remove much of the initial confusion that usually appears when major responsibility for planning and carrying out classroom activities is shifted from teacher to students. As the students became more experienced in group work they did modify some of the suggested responsibilities and added others.

ADVANTAGES OF THE DUAL GROUPING SYSTEM

The forming of permanent work groups and committees, as described above, resulted in a dual grouping system illustrated in the accompanying diagram. The permanent work groups are represented by the six columns and the committees appear in the five rows.

This two-fold chart was used to indicate the permanent work group and the committee to which each student belonged. Since each member of the advisory committee is also chairman of a work group, the names of students on this committee were placed across the top of the chart. The remaining committees were placed down the side of the chart in alphabetical order. The students' names and their major teaching fields were then placed in the squares indicating their respective work group and committee

Permanent Work Groups

Committees	I	II	III	IV	V	VI
Advisory (also chairmen of work groups)	John 1 Smith (Biology)	Don 1 Jackson (History)	Mary 1 Jones (English)	June 1 Hill 2 (Home Ec.)	Bill 1 Brown (Phys. Ed.)	Betty 1 Olson
Evaluation			Martha 2 Black (English)			
Presentation			Jim Johnson (English)		2	
Resource	2		Henry Green (English)			
Visual Aids		2	Ann Morgan (English)			

1—Chairmen of work groups; 2—Chairmen of committees.

appointments, as illustrated on the chart for permanent work group III. The completed chart was duplicated and given to the class members for their orientation.

The advantages of this method of grouping may be obvious from the chart. However, the following unique advantages occurred to the writer during the use of this dual grouping system:

1) *Each student has an opportunity to work with mutual friends as well as to develop new acquaintances.*—Since the permanent work groups are based on sociometric choosing, each student has an opportunity to work with several of his mutual friends during work-group sessions. The committees cut across these sociometric choices, so during committee work the student is provided with an opportunity to develop new acquaintances among other class members.

2) *Every member of the class has an active rôle in assisting with the on-going teaching-learning process.*—This has several advantages. First, no student feels left out in the planning and carrying out of classroom activities since everyone serves on a committee. Second, the committees are able to meet during class time with each class member having a definite responsibility. Third, and probably most important, all students obtain experience in apply-

ing the psychological facts and principles that are learned through individual study and general class discussions. The classroom becomes a learning laboratory where theories are put into practice. The unique contribution of dual grouping, in this regard, is the fact that these activities are not dominated by a few aggressive individuals. This method of grouping spreads the responsibilities among the class members, giving them all an opportunity to participate. Studies, such as that of Lewin,[3] suggest that this active participation is essential if the teacher's on-the-job behavior is to be improved.

3) *Dual grouping provides an efficient multiple communication system.*—In each work group there is a member of each committee, and in each committee there is a member of each work group. Consequently, during work-group sessions there is a committee member available to answer questions related to the service activities of each committee. Likewise, suggestions for improving any of these service activities can be jotted down by the respective committee member and carried back to the next committee meeting. It can readily be seen that this also works in reverse. For example, when the advisory committee has set up tentative plans for classroom activities, each committee member can carry it back to his respective work group for discussion and modification before presenting it to the entire class for approval. This multiple communication system prevents much confusion in both work groups and committees, since each student has contact with a representative of every work group and every committee. The multiple contacts of each student are illustrated in the two-fold chart described previously. It will be noted that Mary Jones meets with the students listed across the chart in her row during meetings of the advisory committee, and she meets with the students listed down the chart in her column during meetings of permanent work group III.

4) *The opportunities for experience in leadership are expanded.*—Ten chairmen are required, six for the permanent work groups and four additional for the committees. With dual grouping the only possible overlap in leadership responsibility is the chairman

---

[3] K. Lewin, *The Relative Effectiveness of a Lecture Method and Method of Group Decision for Changing Food Habits*. Washington, D. C.: Committee on Food Habits, National Research Council, June, 1942. (Mineographed)

of the advisory committee who is already chairman of his work group. The nature of dual grouping forces the remaining chairmenships to be spread among the other class members.

5) *Students have an opportunity to work with class members in their own major teaching fields as well as with class members in other teaching fields.*—The work groups, based on sociometric choosing, are usually fairly homogeneous with respect to the major teaching fields of their members. Thus, in the permanent work groups students benefit from the ideas of classmates with similar interests and similar teaching objectives. Since committees cut across these work groups, students also have an opportunity to share viewpoints and discuss issues with class members in different major teaching fields. This type of small group work is very realistic in terms of the students' future departmental and inter-departmental committee work in the public schools.

6) *Student evaluation of other group members is less biased, since dual grouping makes dual evaluation possible.*—Each student is able to evaluate the performance of other students in both his permanent work group and in his committee. Since the permanent work groups contain a large number of mutual friends, it is expected that the students evaluations will be slightly inflated due to the halo affect. However, the committees cut across sociometric choices and therefore contain relatively few mutual friends. Thus the halo affect is reduced to a minimum, and committee evaluations serve as a check on the evaluations received in permanent work groups.

The above advantages are those which appear unique to the functioning of the dual grouping system. There has been no attempt to include the general advantages of student-centered teaching, except as they relate to this method of grouping. However, it should be apparent that with these, or similar classroom procedures, students are learning more than how to work together in groups and committees; they are learning a realistic approach to the solution of professional problems requiring a knowledge of psychological principles; they are developing skill in planning democratic learning experiences, preparing for them, carrying them out, and evaluating the results; in general, they are obtaining both theory and practice in the application of psychological principles to classroom teaching-learning situations. This not only makes the psychological principles more meaningful to the students, but it also provides them with experiences which later can be utilized in their own teaching. As our public schools shift from a teacher-dominated classroom to student-centered teaching, it appears essential that teachers-in-training be provided with experiences in student-teacher planning. Dual grouping appears to be a systematic procedure for providing these experiences.

STUDENTS' EVALUATION OF THE COURSE

Although no intensive effort was made to measure the specific benefits derived from the use of dual grouping, the students' anonymous evaluations of the course reflect directly and indirectly upon the values of this method. On a four point rating scale, over two-thirds of the students indicated that the following conditions existed 'almost all of the time'; the majority of the remaining students indicated that they existed 'most of the time'; two students, or less, indicated that they existed only 'some of the time'; and no students indicated that they existed 'none of the time'.

1) The course had definite goals.

2) The goals of the course were attained.

3) The class periods were well organized and well used.

4) The class periods were interesting.

5) Class time was not wasted in consideration of questions which had little or no significance.

6) Questions from class members were encouraged and carefully considered.

7) Assignments were clear and carefully given.

8) The course content was functional and related to teaching situtions.

9) Experience in this course has been helpful in leading to more independent and critical thinking.

10) Evaluation of the students' work was related to course goals and was fair.

The above items are only indirectly related to the method of grouping used. However, they do imply that the course did not disintegrate into aimless discussions of plans and procedures for carrying out the classroom activities. Items 3) and 5) probably have the most direct bearing on this point. It is felt that this system of grouping, with its provision for definite student responsibilities, contributed to the organization of class periods and to the prevention of wasted time.

In addition to rating the course on the above points, the students were requested to add any comments that would be helpful in improving the course. They were almost unanimous in suggesting that the permanent work groups and committees be organized earlier, in order to avoid the initial confusion they had experienced in the course. Although it is felt by the instructor that the dual-grouping system contributed to the reduction of this initial confusion, there is no claim that it can, or should, remove it entirely. When students first take over the major responsibility for planning and carrying out their learning experiences a certain amount of confusion is inevitable.

The following student comment, given anonymously at the end of the course, illustrates the problem of adjusting to student-centered teaching: "Because I have been accustomed to teacher-dominated courses, the first few weeks were highly frustrating. However, once adjusted to it I enjoyed the course immensely and feel that for once a course will be of value in my teaching. I do feel that I may have neglected my other courses and spent too much time on this one."

The last part of this reaction reflects the motivation that may result from student-centered teaching. Although time was alloted for committee meetings in class, most of the committees voluntarily met out of class on frequent occasions. It is entirely possible that without these out-of-class meetings the dual-grouping system would have been less effective, due to the amount of class time which would then have been needed for committee work.

In addition to the relatively large amount of class time required for planning and carrying out group responsibilities, there are two other persistent problems related to the course approach described above. One is concerned with the amount of direction given by the teacher. The other is related to how well the basic principles of educational psychology are covered.

The course evaluations revealed a discrepancy in the students' perceptions of the amount of direction provided by the teacher. Typical reactions concerning the lack of direction are indicated in the following students' comments:

"At times we wasted time by being unable to make up our minds. Couldn't you help avoid this?" "Perhaps a little more guidance on the part of the instructor would be helpful. However, this course is such that probably too much guidance is wrong."

In contrast to the above plea for more direction and guidance, one student commented: "This course is operated on a democratic principle but everyone feels that they still have to do what the instructor would like them to do. Perhaps most of us are so steeped in traditional procedures that we cannot appreciate democratic procedures."

This conflict in students' perceptions of the extent to which the teacher dominated the class may be partly due to the varying states of dependency among the students. Some would like more direction and others would prefer more freedom in planning their learning experiences. These contrasting feelings were evident when the permanent work groups and committees were presented with the list of suggested responsibilities. Although it was stressed that these were merely suggestive, some class members wanted more details concerning these responsibilities and how they should be carried out, while others felt the duties were too highly structured and didn't permit enough student initiative. Thus, knowing how much direction to give students and when to introduce that direction is one of the persistent problems of this course approach.

The problem of getting a comprehensive coverage of the basic psychological principles related to teaching-learning situations is closely related to the problem of how much teacher direction should be given. Although one does not get a coverage of all conventional material with a problem approach, one expects to bring out the main psychological principles that underlie effective teaching. Several students raised questions concerning the lack of a comprehensive coverage in this course. The following is a typical comment in this regard: "The problem approach to studying has its advantages, but I have worked on only three specific problems. I wonder if I got a complete picture of the field. That is—there may be many principles, ideas, generalizations, and applications that I did not encounter."

This is a valid criticism since the range of material was not as extensive as that covered in a teacher-dominated course. However, with this type of approach, one attempts to train the student so he will continue independent study on the job. Extensiveness of coverage may be partially sacrificed to attain this goal. In addition, it should be pointed out that there is a tendency on the part of the students to overlook the general significance of their classroom activities unless the application of these experiences is pointed out to them. For example, the general applicability of the principles and procedures of evaluating this course was not recognized until the class members had general class discussions concerning the utilization of these same principles in their own teaching. Getting the students to generalize from their experiences in work groups, committees and class discussions requires the constant attention of the instructor. Without this generalization, the students are unable to recognize that most of the material in a beginning educational psychology course is being covered, but in a different way. The following student comment reflects a sudden realization that content was fairly well covered in the course: (The test referred to is a comprehensive multiple-choice test that was responded to out of class, at the end of the course.) "I didn't realize how much I had gained from the professional records, panel discussions, group work, and class discussions until I was looking over the take-home test. I then realized we had discussed many of the concepts in the test."

The student evaluation material presented above is not offered as proof that this is the best approach to student-centered teaching. It is merely descriptive of the various student reactions to one instructor's attempt to use a systematic grouping method to facilitate student-teacher planning. It is hoped that these student reactions will provide additional insights for any instructor who might experiment with dual grouping in his own teaching.

SOLUTIONS BY TEAMS AND BY INDIVIDUALS TO A FIELD PROBLEM AT DIFFERENT LEVELS OF REALITY[1]

IRVING LORGE, JACOB TUCKMAN, LOUIS AIKMAN,
JOSEPH SPIEGEL and GILDA MOSS

Teachers College, Columbia University

In teaching and in training, problems are presented most often verbally, usually as a printed text. Despite the preponderance of practice employing verbal presentation, advanced instructional thinking seems to favor teaching, and hence, learning by realistic modes of presentation. By implication, the order of preference in teaching modes would range, let us say, from the most frequently advocated 'actual real' through the three-dimensional scale model, the two-dimensional photographic representation, to the least recommended 'verbal abstraction'. Such modes of presentation are assumed to be a range of steps or levels differing in remoteness from reality. Of course, additional levels of remoteness may be specified. For instance, in learning to pilot an airplane, other methods of presentation may be employed such as: textbook exposition, diagrammatic representation, exploded assembly diagrams, simulated trainers and actual aircraft.

The purpose of this study is to estimate the difference in the quality of the solution to a practical field problem presented in four settings differing in their degree of remoteness from reality: a) the verbal description, b) the photographic representation, c) the miniature scale model, but not allowing manipulation of parts and materials, d) the miniature scale model allowing manipulation of parts and materials.

In order to estimate the relative effectiveness of different levels of presentation for problem solving, it was considered desirable to find or to devise a problem which (1) could be adapted for presentation at different levels of remoteness from reality, (2) would permit many different solutions of varying quality not only of thought, but of accomplishment, (3) would have the characteristics of a genuine field situation, usually lacking in most psychological laboratory problems and (4) could be used with groups as well as individuals.

---

[1] *Studies in Reality*. No. 1. Conducted under AF 18 (600) 341 of the Human Resources Research Institute, Maxwell Air Force Base, by the Institute of Psychological Research, Teachers College, Columbia University, Principal Investigator: Irving Lorge.

The field problems developed by the Office of Strategic Services during World War II to assess leadership and personal qualities of men appeared to satisfy these criteria. 'The Mined Road Problem'[2] which requires the formulation of a plan for getting a group of five men across a road mined with supersensitive mines which cannot be neutralized or dug up was selected and adapted.

The Mined Road Problem was constructed in reality in wooded terrain. The section of the road which was to be crossed was twelve feet wide and approximately fifty feet long. The limits within which the subjects were permitted to work extended about fifteen feet away from the road on either side. Two-inch tape marked off the road as well as the working limits. On each side of the road were two trees about forty feet high: two were directly opposite one another and just at the edge of the road. Scattered about the work area on the starting side were the following materials which could be utilized in crossing:

2 4″ x 6″ wooden beams, one twelve feet long, the other fourteen feet long
1 8″ x 8″ wooden block, thirty inches long
3 2″ x 4″ boards, four feet long
2 1″ x 4″ boards, eighteen inches long
1 automobile tire
2 ½″ Manila ropes, one twenty-five feet long, the other fifty feet long
1 pulley
1 hammer

After the Mined Road Problem had been constructed in the actual field setting, four methods of presentation were developed:

(1) *The Verbal Description.*—The problem was presented in typed form to the men[3] as follows:

> You men are on your way back from a sabotaging mission in enemy-occupied territory. You have just blown up a bridge about a mile from here. According to a prearranged plan, you will meet a guerrilla truck about a mile away from here which leaves you only a very short time to get across this road. You have discovered that this road has been mined with a new

---

[2] The OSS Assessment Staff. *Assessment of Men.* New York: Rinehart & Co., Inc., 1948.

[3] The OSS Assessment Staff. op. cit., p. 148.

type of enemy mine which is supersensitive and will blow up if anything touches it. The explosion at the bridge has aroused the enemy but as yet they do not know in what direction you are going. Your problem is to get the entire group of five men across the road and leave as little trace as possible of your escape route.

V Bordering the road are trees and scattered around are some wooden beams, some pieces of lumber, some lengths of rope, a few sticks, an old tire, a pulley, a hammer, and some other debris.

After you have decided how you are going to cross the road, one man is to write your plan of action on the paper provided.

You may ask any questions you wish and we will try to answer them as best we can. Remember it is important that you complete this mission as quickly and with as little loss of men and materials as possible. According to your schedule you must meet the guerrilla truck in one hour. Remember also that your performance on this mission will be compared with other members of Air Force Reserve Officers Training Corps personnel.

(2) *The Photographic Representation.*—This presentation consisted of five photographs of the Mined Road Problem taken from five different positions on the original starting side. Each photograph enlarged to 8″ x 10″ was numbered to correspond to the angle from which it was taken. The photographs were bound in a booklet together with a schematic diagram, indicating the positions relative to the terrain from which each photograph was taken. The typed description of the problem presented to the men was identical with that of the Verbal Description except that the following paragraph was substituted for paragraph *V*:

You must cross the road at the spot on the sketch labeled 'mission's present position'. You have five photographs of this mined road taken from five different angles. Photograph no. 1 shows the road from position no. 1 on the sketch. Photographs nos. 2, 3, 4, and 5 show the road from positions 2, 3, 4, and 5 respectively, on the sketch. The white tape in the photographs corresponds to the lines in the sketch indicating the road and working limits.

(3) *The Miniature Scale Model Without Manipulation.*—A miniature scale model of the Mined Road Problem, built to the scale of one inch equals one foot with over-all dimensions about three feet by three feet and with the trees approximately three feet in height, duplicated the actual field situation, i.e., scale representations of beams, crosspieces, tire, ropes, etc. were distributed about the model as in the field.

The typed description of the problem given to the men was identical with that of the Verbal Description except that the following paragraph was substituted for V:

> You must cross the road at this spot which is marked off by the white tape. In determining your plan of action, you may observe but not manipulate any parts of the model or equipment.

(4) *The Miniature Scale Model With Manipulation.*—The same model was used. The essential difference between No. 3 and No. 4 is that the men could manipulate the beams, tire, ropes, etc. The typed description of the problem given to the men was identical with that of the Verbal Description except that the following paragraph was substituted for V:

> You must cross the road at this spot—which is marked off by the white tape. In determining your plan of action you may manipulate any parts of the model or equipment.

The subjects were cadets of the Air Force Reserve Officers Training Corps at Columbia College, Columbia University, in Air Science I and Air Science II. These men were fairly homogeneous in background and intelligence, were highly motivated to work with tactical problems, and would maintain the required security about the problem.

Ten teams of five men and ten men as individuals were set the task of solving the problem at each of the four levels of presentation. If the problem was to be presented for team solution, the five team members were brought as a group to the laboratory and given one of the typed descriptions of the problem to read to themselves while the examiner read it aloud. If the problem was to be presented for individual solution, only that person was brought to the laboratory, given the typed description of the problem to read himself while the examiner read it aloud.

Since the information available about specifications of the road, the terrain, and the materials varied from level to level, i.e., according to the form of presentation by the very nature of the presentation method, teams and individuals were permitted and encouraged to ask questions in an attempt to equalize the amounts of information from team to team, from individual to individual, and, more important, among presentations. When an individual or a team asked a question about information that could have been obtained by observation in the actual field setting, e.g., the width of the road, length of the beams, height of the trees, thickness of the rope, such question was answered in required detail.

When the question was not answerable by observation in the actual field setting, e.g., location of the mines, types of mines, ability of men to swing across the road or to broadjump across the road, the examiner stated that he did not have the information. The teams and individuals proceeded with their interrogations and deliberations until the solution was ready to be written. Then either the individual or a member of the team wrote the solution. Thus, eighty written solutions were obtained: ten by teams and ten by individuals for each of the four methods of presentation.

The written solutions of the Mined Road Problem do not represent just one correct solution, but rather they illustrate a rich variety differing not only in the degree of workability or accomplishment, but also varying in the quality of thinking as well. Obviously, a 'go-no go' scoring would be inadequate. To develop a scoring procedure that would credit the quality of solutions, a content analysis of the solutions was made. All solutions were oriented around four major elements: (1) Bridging the road, i.e., the use of beams, a combination of beams, the swinging, broadjumping, throwing men or materials across, making an overhead bridge, etc. (2) Removing the bridging from the road, i.e., any procedure for removing the basic bridging or any secondary bridging. (3) Removing clues indicative of escape or escape route, i.e., materials hidden or carried away. (4) Time, i.e., elapsed time from presentation of problem to the formal written solution. Each point and subpoint in all written solutions was classified by a detailed content analysis. To give a score, each item in the content analysis was assigned a weight or value by giving consideration to the aggregate of factors such as safety, efficiency and quality of thinking. In assigning weights, three members of the research team tested the workability of various solutions in the actual real setting. For instance, it was found by tryout that solutions involving a man swinging by rope from a tree on the starting side would not work except under the joint specific conditions that the rope had to be suspended from a branch at least twenty feet high, and at the same time the man must begin the swing from an elevation of about six feet from someone else's shoulder or from a piece of lumber to get the necessary momentum. Or, it was found that a solution involving a man broadjumping across the road would not work at all since the run space was only fifteen feet long and on an upgrade. The Physical Education

11.8.70
9103

Department of Columbia College stated that it was highly improbable that anyone could broadjump the road under the actual conditions especially if the groups were out of condition or if the broadjumper lacked necessary track equipment. On the other hand, solutions involving throwing a tire with attached rope across to the other side of the road would work even though there was some danger that the rope or tire might hit the road.

After the workability of the solutions had been field-tested, values were assigned to various aspects of the written solutions. For example, in bridging the road with a beam or combination of beams, the fourteen-foot beam was given a weight of 25, whereas the fourteen-foot beam tied to the twelve-foot beam was given a weight of 20; or, to give another instance, the twelve-foot beam tied to the fourteen-foot beam about one foot from its end, spread into an X, was given a weight of 15. The suggested use of a beam without specifying its length was given a weight of 8; whereas the use of the twelve-foot beam (which could not bridge the road) was given a weight of 0. Similarly, removing or leaving clues indicative of the escape route were given differential credits: for example, leaving all materials on the original starting side and removing footprints was given a weight of 25; leaving some materials on the starting side and hiding some materials on the other side and removing footprints from both sides of the road was given a weight of 20; leaving some materials on the starting side of the road and some on the other side and removing footprints was given a weight of 15; no reference to removal of materials on either side of the road, yet implying removal, was given a weight of 7; leaving a rope suspended across the road between two trees was given a negative weight of −8; leaving a beam bridging the road as well as leaving a rope suspended across the road between two trees was given a negative weight of −15.

After the Quality Point Scoring system had been developed, fifteen written solutions selected at random were ranked for over-all workability and goodness of solution by three members of the research team. Independently, the Quality Point Score for each of the fifteen solutions was obtained by summing the separate element weights. The correlation between Quality Point Score and rank for workability approximated unity. The Quality Point Score seems an objective way of appraising goodness of solution.

TABLE 1.—QUALITY POINT SCORE FOR MINED ROAD SOLUTIONS: MEANS AND STANDARD DEVIATIONS, BY TEAMS AND BY INDIVIDUALS AT EACH OF THE FOUR LEVELS OF REMOTENESS FROM REALITY: TEN TEAMS AND TEN INDIVIDUALS AT EACH LEVEL

Level of Remoteness from Reality	Teams		Individuals	
	Mean	SD	Mean	SD
Verbal description	68.9	28.2	25.2	16.8
Photographic representation	54.8	18.3	11.7	9.2
Miniature scale model without manipulation	50.1	24.6	34.0	27.4
Miniature scale model with manipulation	59.0	19.7	36.7	26.3

TABLE 2.—QUALITY POINT SCORE FOR MINED ROAD SOLUTION: ANALYSIS OF VARIANCE BETWEEN TEAMS AND INDIVIDUALS AND AMONG FOUR LEVELS OF REMOTENESS FROM REALITY

Source	Sum of Squares	Degrees of Freedom	Mean Square	F Ratio
Class (teams and individuals)	19593.8	1	19593.8	35.7*
Treatment (level of reality)	2701.6	3	900.5	1.6
Class X treatment	3025.2	3	1008.4	1.8
Error	39367.2	72	546.8	
Total	64687.8	79	818.8	

* Significant at .01 level.

For teams, and for individuals, at each of the four levels of remoteness from reality, the means and standard deviations of the Quality Point Scores are given in Table 1. The analysis of variance among levels of remoteness from reality and between teams and individuals is presented in Table 2.

The evidence indicates that there are no significant differences among the solutions at the four levels of remoteness from reality. When teams solved the Mined Road Problem, the goodness of the written solutions was approximately equivalent for the Verbal Description, Photographic Representation, Miniature Scale Model Without Manipulation, and Miniature Scale Model With Manipulation. When individuals solved the problem, their written solutions were approximately equivalent at every level of remoteness from reality. Apparently the written solutions for the four levels cannot be distinguished in terms of quality. But the evidence does show that at every level of remoteness from reality, the written solutions by teams is markedly superior to those by individuals.

The absence of differences in goodness of solution to the Mined Road Problem among levels of presentation may be attributed, at least in part, to the fact that teams and individuals at each level were given the opportunity to obtain equal amounts of information about the problem and its materials either by perception directly or indirectly by asking questions. This suggests that many of the so-called differences among presentation modes may really represent differences in the amount and kind of information the mode allows. Adjustment for presentation mode does lead to the formulation of equally good hypotheses and, hence, equally good solutions for the problem. The superiority of team solutions over individual solutions in part may be due to the fact that teams of five members ask more questions than do single individuals. Teams, thus, not only get more information, but also seem to evolve many more fruitful hypotheses regarding a solution. The team thus summates information and hypotheses leading to better solutions. This experiment corroborates the effectiveness of groups in the so-called 'Eureka' problems—problems with workable and testable solutions.

A FACTOR ANALYSIS OF THE SRA YOUTH INVENTORY

JOHN P. PAISIOS and H. H. REMMERS

Division of Educational Reference, Purdue University

The *SRA Youth Inventory* (*1*, *2*) is a needs and problems inventory that has been specifically developed to help teachers, counselors, and school administrators quickly identify the problems of their high-school youth. The items of the Inventory were developed primarily from a content analysis of hundreds of anonymous essays written by a national sample of teen-agers. The items finally selected were divided *a priori* into eight areas called (1) My School, (2) Looking Ahead, (3) About Myself, (4) Getting Along with Others, (5) My Home and Family, (6) Boy Meets Girl, (7) Health, and (8) Things in General. The preliminary form of the Inventory was administered to approximately fifteen thousand students from high schools participating in the Purdue Opinion Panel.

Using a sample of one thousand students drawn from the national sample, the homogeneity of each problem category was established by means of an item analysis. A biserial coefficient of correlation was computed between each item and the total score in each problem area. The item analysis data were used to reassign items from one category to another in order to increase the homogeneity and, hence, the psychological meaning of each area.

Intercorrelations of the total scores for the eight 'purified' problem categories were computed for the sample of one thousand used in the item analysis. Reliabilities for each of the problem areas were also computed, using the Kuder-Richardson Formula, Case II. These intercorrelations and reliabilities are presented in Table 1.

PURPOSE

Although the eight test areas, by virtue of the item analysis, are relatively independent and possess unique variance, they do have variance in common. The entire matrix is a positive manifold. Seven of the twenty-eight intercorrelations are .50 and larger. Any two of these highly intercorrelated tests have a common variance of at least twenty-five per cent. Accordingly, it was the purpose of this study to analyze statistically the overlap of vari-

TABLE 1.—RELIABILITIES AND INTERCORRELATIONS OF TOTAL SCORES FOR THE EIGHT PROBLEM AREAS
(N = 1000)

Area	1	2	3	4	5	6	7	8
1. My School	(.84)	.40	.55	.45	.42	.35	.34	.20
2. Looking Ahead		(.90)	.46	.58	.33	.46	.36	.39
3. About Myself			(.88)	.67	.50	.45	.47	.35
4. Getting Along with Others				(.88)	.46	.59	.51	.48
5. My Home and Family					(.94)	.46	.47	.32
6. Boy Meets Girl						(.87)	.50	.48
7. Health							(.75)	.46
8. Things in General								(.89)

ance in an attempt to identify psychologically meaningful problem factors operating in the eight areas of the Inventory.

PROCEDURE

The matrix of intercorrelations was factor analyzed using Thurstone's centroid method (*3*). Since the reliabilities of the subtests were known, it was possible, once the communalities were determined, to account for the total variance of each subtest (*3*). That part of the variance which is unique was factorially determined and the unique variance was separated into its two parts, the specificity and the error variance.

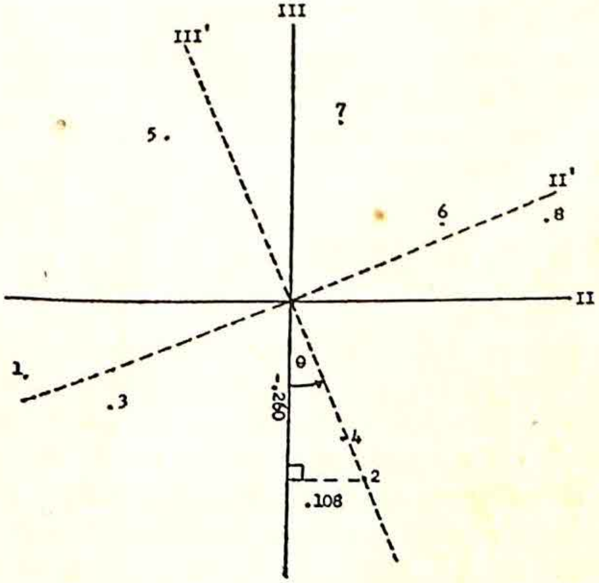
RESULTS

Three factors were extracted from the intercorrelation matrix. The original and rotated factor loadings appear in Table 2. The rotation scheme adopted was a simple counterclockwise rotation in the plane II–III such that $\theta = \text{arc sin } .3837$. The projection of the eight test vectors upon the plane II–III before and after rotation appears in Figure 1. It will be noted from Figure 1 and Table 2 that seven of the new loadings in Factors II and III are zero or nearly so. Every subtest has one zero loading in either Factor II or III except Health, which is positively loaded in both. It will also be noted that the rotation has established two bipolar factors (II and III) and a principal factor (I). This rotation system produced the most meaningful structure. Rotation systems were attempted which resulted in a positive manifold of factor loadings but these structures proved to be relatively meaningless.

TABLE 2.—ORIGINAL AND ROTATED FACTOR LOADINGS, WITH THE COMMUNALITIES, UNIQUENESS, SPECIFICITY, AND ERROR FOR THE EIGHT PROBLEM AREAS

Area	Original Factor Loadings			Rotated Factor Loadings			Proportions of Variance			
	k_{I}	k_{II}	k_{III}	k_{I}	k_{II}	k_{III}	h_j^2	u_j^2	b_j^2	c_j^2
1. My School	+.601	−.370	−.113	+.601	−.385	+.038	.511	.489	.329	.160
2. Looking Ahead	+.656	+.108	−.260	+.656	.000	−.282	.510	.490	.390	.100
3. About Myself	+.759	−.251	−.155	+.759	−.291	−.047	.663	.337	.217	.120
4. Getting Along with Others	+.812	+.079	−.199	+.812	−.003	−.214	.705	.295	.175	.120
5. My Home and Family	+.637	−.176	+.240	+.637	−.070	+.289	.494	.506	[illegible]	.060
6. Boy Meets Girl	+.715	+.213	+.110	+.715	+.239	+.020	.569	.431	.301	.130
7. Health	+.667	+.070	+.259	+.667	+.164	+.212	.517	.483	.233	.250
8. Things in General	+.582	+.364	+.112	+.582	+.379	−.036	.484	.516	.406	.110

Projection of the Eight Test Vectors[1] upon the Plane II-III, with a Rotation through θ = arc sin .3837 Indicated



[1]Vector Code	Area
1	My School
2	Looking Ahead
3	About Myself
4	Getting Along with Others
5	My Home and Family
6	Boy Meets Girl
7	Health
8	Things in General

FIG. 1

Table 2 also presents an account of the total variance of each subtest. The total variance has been fractionated into that which is common (h_j^2) and unique (u_j^2) and the unique variance has been broken down into those parts due to specific elements (b_j^2) and to errors of measurement (e_j^2).

DISCUSSION

Interpretation of the Factors.—The naming of Factor I is seemingly obvious since it makes up most of the variance in every problem area. Because of its high loadings in every subtest, this factor is interpretated as a 'General Problems' factor.

Factors II and III proved to be more difficult for extracting a general naming principle. The rotation solution has resulted in two high positive and two high negative loadings in Factor II (see Figure 1 and Table 2). We have a continuum, with Things in General and Boy Meets Girl opposed to My School and About Myself. At first glance, these area titles do not seem to lend themselves to a higher order abstraction. Examination of the problem items per se, however, suggests a scope continuum; that is, the problems expressed seem to progress from those of an internalized, personal nature (My School, About Myself) to those of an externalized, social nature (Boy Meets Girl, Things in General). A few items from each area are reproduced below to illustrate the nature of the continuum. The problems are arranged to progress from the internal to the external pole of the continuum.

Area	*Item*
1) My School	I have difficulty keeping my mind on my studies. I am too restless to stay in school.
2) About Myself	I prefer to be alone, I wonder if I am normal in the way my mind works.
3) Boy Meets Girl	What are good manners on a date? I need more information about sex.
4) Things in General	How can we get honest government? How can I help get rid of intolerance?

Factor II, therefore, in view of the above axis, can be given the hyphenated title of 'Internal-External' or perhaps 'Egoistic-Social' factor.

Loadings in Factor III set up My Home and Family at one end of a continuum and Looking Ahead and Getting Along with Others at the opposite end. The problems in the above areas suggest a time continuum; that is, problems included in My Home and Family seem to have historical and present-time stress, whereas problems included in Getting Along with Others and Looking Ahead generally place emphasis on the future. Accordingly, Factor III can be looked upon as a 'Historical-Future' factor. Selected problems from the areas under discussion are presented below. The problems range from the historical to the future axis.

Area	*Item*
1) My Home and Family	My parents neglect me. My family tries to hold me down.
2) Getting Along with Others	I want people to like me better. I want to make new friends.
3) Looking Ahead	What career shall I pursue? What are some ways of financing a college education?

The Specific Variance.—The contribution to the total variance of each subtest by the three common factors has now been extracted and identified. Examination of Table 2, however, reveals that considerable specific variance remains to each area. The eight problem areas retain their psychological meaning because of their possession of specific elements above and beyond errors of measurement. The identification of three underlying factors has not destroyed the original problem areas of the Inventory, but has, however, provided more information about the problems confronting teen-agers.

SUMMARY AND CONCLUSIONS

The *SRA Youth Inventory*, its definition, purpose, and construction, was briefly discussed. Special emphasis was placed on the intercorrelation of the eight problem areas and the desirability of a factor analysis was pointed out. The matrix was factor analyzed using Thurstone's centroid method. With the communalities determined, and with the reliabilities known, the total variance for each problem area was accounted for.

The following conclusions are supported:

1) Three underlying factors are operative in the problem areas of the *SRA Youth Inventory.*

2) The factor which contributes most to the variance is identified as a 'General Problems' factor because of its high loadings in every subtest.

3) A second factor is identified as an 'Egoistic-Social' factor.

4) The third factor is called a 'Historical-Future' factor because it suggests a time axis.

5) The eight problem areas contain substantial specific variance after the extraction of the variance due to the common factors and to the errors of measurement.

6) The fractionization of the variance common to all the areas into three meaningful problem factors, coupled with the knowledge of the factors specific to each area, provides more information on the nature of the needs and problems of younger people.

REFERENCES

1) H. H. Remmers, B. Shimberg, and A. J. Drucker. *Technical Supplement for the SRA Youth Inventory, Form A.* Chicago: Science Research Associates, 1953.

2) B. Shimberg. "The development of a needs and problems inventory for high-school youth." *Studies in Higher Education* LXXII; *Further studies in attitudes, Series* XVII. Lafayette, Ind.: Division of Educational Reference, Purdue University, 1947.

3) L. L. Thurstone. *Multiple Factor Analysis.* Chicago: University of Chicago Press, 1947.

THE USE OF AN ELECTRONIC COMPUTER IN PRINCIPAL AXES FACTOR ANALYSIS[1]

CHARLES WRIGLEY and JACK O. NEUHAUS[2]

University of Illinois

Psychology promises to benefit greatly from the development of electronic computers. Most of the day-to-day calculations in psychometrics can be carried out at greater speed and with a higher degree of accuracy on an electronic computer than by more conventional computational techniques. It follows that exact solutions become practicable in situations where psychologists have hitherto been restricted to approximate methods. On the other hand, the transitional period in the changeover to electronic computer procedures is not easy. Programming for an electronic computer is difficult and requires special training, and new skills have to be developed in the recording and handling of tapes.

This report has been prepared to consider the use of the electronic computer for principal axes factor analysis (*9*). The advantages of the principal axes method may be summarized thus:

1) Vectors of factor loadings are orthogonal. In the centroid method, there is a loss of orthogonality, i.e., scalar products of columns of factor loadings are not exactly equal to zero.

2) A maximum of the variance is accounted for by each successive factor; the principal axes solution is thus unique (except where there are multiple latent roots).

3) Because the matrix of latent vectors from which the factor

---

[1] This research was supported in part by the United States Air Force under Contract No. AF 33(038)-25726, monitored by the Human Resources Research Center. Permission is granted for reproduction, translation, publication, use and disposal in whole and in part by or for the United States Government.

[2] We wish to express our deep gratitude to Dr. R. E. Meagher, Director of the Digital Computer Laboratory, and to Dr. J. P. Nash, senior mathematician of the Laboratory, for the way in which they have placed their facilities at our disposal, and for their generous help; and to Mr. R. T. Gregory for preparing the code to give the latent roots and latent vectors of a matrix, from which our program for principal axes factor loadings was derived.

loadings are derived is orthonormal, the inverse of the matrix of factor loadings may be readily computed. This makes it somewhat easier to get the factor measurements.

4) Tests of significance have been developed for the successive latent roots (i.e., the sums of squares of loadings in successive factors) if unities are inserted in the principal diagonal of the correlation matrix (*2*, *3*).

The difficulty hitherto has lain in computing the latent roots and latent vectors from which the principal axes factor loadings are derived. As a result, very few principal axes solutions are available in the psychological literature. The largest correlation matrix handled by the method appears to have been one of thirty-one variables. This was calculated by Deemer, and has not been published.* Davis (*6*, *7*) has worked with matrices of order 14 and order 9, and Kelley (*12*) has handled one of order 10 and another of order 9. We understand that sixty hours' computing was required for a matrix of order 7 when results were needed with four-place accuracy (*5*), and Holzinger and Harman (*8*, p. 175) report that one hundred ten hours were needed to extract the first two factors for a matrix of order 24.

The centroid method of factor analysis has generally been used by psychologists in preference to the principal axes method. The advantage of the centroid method has been that it has given a reasonably good approximation to a principal axes factor analysis in only a small part of the time otherwise required. The mathematically superior principal axes method has demanded too much arithmetic. With an electronic computer, however, it is quite simple to get the principal axes factor loadings once the code has been written. The factor loadings for a correlation matrix of order 7 are computed and punched on tape in about a minute, with results to about ten decimal places; a matrix of order 20 is analyzed in about ten minutes, giving the entire set of twenty factors.[3] Even allowing ten or even twenty times as long as this for preparation of tapes and checking of results, this is much quicker than current methods of computation. Hence the principal axes method is to be preferred to the centroid as the standard starting point in factor analysis whenever an electronic computer is available. When rotation of axes is decided upon, it will be the principal axes solution which is transformed.

---

* F. B. Davis. Personal communication March, 1954.

[3] Two further examples may be given of the shortened times for computation using the Illiac. The full set of product-moment correlations between forty tests can be calculated and results punched on tape in about fifteen minutes. Again, a set of equations involving about forty unknowns can be solved in about five minutes.

THE ELECTRONIC COMPUTER

Several years ago the University of Illinois established a research unit for work in high-speed computing. The Ordvac, built on contract for Aberdeen Proving Ground, Maryland, was completed towards the end of 1951. A rather similar machine, the Illiac, was completed in the latter part of 1952, and is available for the University's own use.

The principal features of the two machines are:

1) *The high speed of operation.*—The Illiac receives its orders and problem data from teletype tapes run through a high-speed photo-electric reader at the rate of about one thousand eleven-place decimal numbers per minute. Once the input of orders and data is complete, the calculations begin. The Illiac averages about one thousand arithmetical operations per second, where by arithmetical operation is meant a completed addition, subtraction, multiplication or division. As soon as the calculations are completed, the results are punched out as decimal numbers on a tape at the rate of about one hundred eleven-place numbers per minute.

2) *The high degree of accuracy.*—The Illiac gives results to as many as eleven decimal places if so ordered. When allowance is made for (a) rounding-off errors in the final position and (b) the scaling involving the introduction of a leading zero, results are generally accurate to at least nine decimal places.

3) *The rather limited storage capacity at present.*—In all, there are only 1024 positions in the 'memory', and some two hundred of these must be allocated to the orders for the principal axes factor loadings, so that only about eight hundred numbers may be retained by the 'memory' at any one time. It follows that there is a restriction at present on the size of matrix that may be analyzed. A correlation matrix of order 23 is the largest for which a full set of principal axes factor loadings can be obtained. A somewhat larger matrix can be diagonalized by making two or more analyses in succession, and printing a selection of factor loadings on each occasion; we ourselves have handled a matrix of order 30

in this way.[4] The Digital Computer Laboratory intends to add a magnetic storage drum to the Illiac, and when this is done there will be a great increase in storage space.

4) *The complexity of the coding.*—A scheme of numerical analysis is required in a form that may be handled by the machine, and then orders have to be written in the form of code-numbers for each stage of the arithmetic. The coding is the most difficult operation in the use of the machine. The more skilled the programming, the more succinctly orders may be expressed and, accordingly, the more space is available for the numbers which are being operated upon. There are usually alternative methods for any standard statistical operation. The programmer has to select a method which (a) is easily coded, (b) is economical in its use of storage space, and (c) involves proportionately fewer arithmetical operations. (In the usual type of psychological calculation customarily carried out by more conventional methods, this third proviso is not important, since computing time is generally negligible, but it is important for those types of analysis which require extensive electronic computer time—fifteen minutes or more of electronic computer time for calculating, say—because machine time is valuable.)

On a desk calculator, the principal axes loadings for the first factor are obtained by the multiplications $R^{(0)} \cdot a^{(1)}$, $R^{(0)} \cdot a^{(2)}$, etc., where $R^{(0)}$ = the correlation matrix and $a^{(1)}$, $a^{(2)}$, etc., denote trial vectors, until the values of the successive trial vectors $a^{(i)}$, $a^{(i+1)}$ are stationary or differ by less than any specified amount.

---

[4] Factor analyses have been made of as many as two hundred variables on the Illiac, using the square root method of factor analysis, where the computational procedure may be arranged so that only one column of the correlation matrix is operated upon at any one time. (See: C. F. Wrigley and L. L. McQuitty, *The Square Root Method of Factor Analysis: A Re-examination and a Shortened Procedure.* Manuscript on file in the University of Illinois Library, 1953). Let the square root factor loadings be denoted by G. Then, by solving the equation $(G'G - \lambda I)\, v = 0$, where v denotes the latent vectors of G′G, we may find the orthogonal transformation matrix V such that $GV = \overline{F}$, where $\overline{F}$ is orthogonal. The factor loadings $\overline{F}$ are then equal to the principal axes factor loadings of $\overline{R}$, where $\overline{R} = GG'$. By taking sufficiently many factors, $\overline{R}$ may be made to approximate increasingly closely to R, so that it is appropriate to speak of this procedure as giving an indirect principal axes factor analysis of R. By means of this procedure, the Illiac is restricted to approximately thirty factors instead of to approximately thirty variables.

After convergence, the trial vector is proportional to the first latent vector, and the largest element in the trial vector is equal to the first latent root. Then $f_{i1} = u_{i1} \cdot \sqrt{\lambda_1}$. The correlation matrix $R^{(0)}$ is reduced to the first residual matrix $R^{(1)}$, and the operation is repeated. Various modifications have been introduced to systematize computational procedure and accelerate rate of convergence (*1*, *4*, *10*).

This procedure designed for desk calculator would be clumsy on an electronic computer. The present Illiac program provides for the diagonalization of the correlation matrix by sine-cosine transformations to reduce each off-diagonal element to zero. For example, in diagonalizing a correlation matrix of order 3

$$\begin{bmatrix} r_{11} & & \\ r_{21} & r_{22} & \\ r_{31} & r_{32} & r_{33} \end{bmatrix},$$

where the matrix is symmetrical so that we record only the lower left-hand portion, the electronic computer first makes the sine-cosine transformation which reduces r_{21} to zero, by multiplying R by the orthogonal transformation matrix.

$$\begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The appropriate angle θ and its sine and cosine are determined by the machine. The electronic computer then makes a second sine-cosine transformation so that r_{31} becomes zero, and then a third so that r_{32} becomes zero. This represents one complete iteration. By this time, however, r_{21} and r_{31} are no longer zero. Nonetheless, it may be shown that this procedure gives convergence, so that after a number of iterations all non-diagonal elements are zero. In another part of the 'memory', an identity matrix of order 3 is stored which records the net transformation, and the resultant is the matrix of latent vectors, which is then converted by the machine to the matrix of factor loadings. The elements remaining in the principal diagonal of R after its diagonalization are the latent roots.

This sine-cosine procedure for the diagonalization of a matrix was originally developed by Jacobi (*11*). The method was presented with explicit reference to the calculation of principal axes factor loadings by Kelley (*12*) in 1935.

With the correlation matrices which can be analyzed with the present memory capacity, this method works very satisfactorily, but since the time required for convergence increases approximately according to the cube of the order of the matrix—e.g., an order 30 matrix requires approximately eight times as long to converge as an order 15 matrix—the analyses of large matrices theoretically made possible by the addition of the storage drum would require so much calculational time that they would be impractical. A code is planned using a method devised by Lanczos (*23*) which, while more elaborate, nonetheless requires fewer operations and is expected to cut computational time to about one-fifth that of the sine-cosine method.

The difficulty of coding even the present relatively simple mathematical procedure becomes evident when we consider that an electronic computer can use only a few simple types of arithmetical operations in carrying out its computations. During the course of the solution of a principal axes problem, the machine selects elements from a matrix, generates sine-cosine matrices and square roots, makes linear transformations, converts the latent vector elements to factor loadings, and punches results on teletype tape. Orders have to be written to carry out all these operations in the appropriate sequence.

Further, each relay in the machine has only two positions—a unit position and a zero position. Since this feature makes use of the decimal system impossible, the machine uses binary numbers which, for coding convenience, are themselves grouped into sexadecimals. The first portion of the order code provides for transformation of decimal numbers to binary numbers during the input, and there is provision at the end of the order code for the converse transformation at the time of output of the results.

Some months' work was needed for preparing the code.[5] However, this one set of orders may be used for all principal axes factor analyses within the storage capacity of the machine. The only tasks now are to prepare the teletype tapes containing these orders and correlations and to print out and check the results.

---

[5] For the reports upon this code, see: R. T. Gregory. *Eigenvalues and Eigenvectors of a Symmetric Matrix.* University of Illinois, University Research Board Electronic Digital Computer Report No. 42. See also: J. O. Neuhaus, *Conversion of Latent Vector Elements to Factor Loadings.* (Manuscript on file in the University of Illinois Library).

The orders may be used with any symmetric matrix. There is no restriction to product moment correlations; the factor loadings may be obtained for any measures of correlation; e.g., tetrachoric correlations, phi coefficients, or indeed for any measures of association. If (a) the measurement matrix is multiplied by its transpose during the computation of the indices, as happens in getting covariances, product moment correlations, and phi coefficients, and (b) the diagonal entries are not reduced in any way, then the matrix may be shown to be positive definite (or positive semidefinite, if its rank is less than its order).[6] Since the matrix is also symmetrical, it is said to be Gramian.[7] If the matrix is positive definite, all latent roots are positive. If the matrix is positive semidefinite, there will be $n - m$ zero latent roots, where n = the order of the matrix and m = the rank.

Analyses can be carried out equally well with communalities inserted in the diagonal in place of unities. (Variances might equally well be reduced in the factor analysis of a variance-covariance matrix with the purpose of achieving minimal rank for the matrix.) However, if any communalities are underestimated, then the matrix will no longer be Gramian. Some of the latent roots will accordingly be negative and the associated factor loadings will be imaginary numbers. In these circumstances the Illiac supplies the correct latent root. The machine proceeds to calculate the square root of the absolute value of the latent root, so that the factor loadings are punched as real numbers. It is then necessary to supply only the symbol i at the appropriate places on the printed copy. However, all the main components will have positive roots, and any factor with a negative root may be disregarded in examining results.

If the investigator wishes to insert communalities instead of unities in the diagonal, his best procedure is to analyze initially with unities. If he then determines the number of factors he wishes to retain, he can estimate communalities (the new teletype tape is readily prepared by copying the original, changing only the diagonal entries), re-analyze, and if needs be repeat the procedure on several occasions. If the factor analysis is made on product moment correlations with unities in the diagonal, then Bartlett's test (*2, 3*) may be used to determine the number of significant components.

---

[6] This means that all the principal minors of the matrix are greater than or equal to zero.

[7] L. R. Ford, Jr. *A Proof That the Matrices of Certain Indices of Association Are Gramian.* (Manuscript on file in the University of Illinois Library, 1953).

THE COMPUTATIONAL PROCEDURE

Two teletype tapes are read into the machine:

1) *The order tape.*—These are the general orders for the calculation of a principal axes factor analysis, applicable to any order of matrix within the storage capacity of the Illiac. These orders provide for the generation of the identity matrix of order n by the electronic computer for storage in the 'memory'. The tape is held on file, and may be used in a series of analyses.

2) *The problem and parameter tape.*—This tape consists of two parts, the problem data and the parameters. The problem section gives the correlations to be analyzed. Since the correlation matrix is symmetrical, only the principal diagonal and the lower left-hand portion of the matrix are punched on tape. The parameter section indicates to the machine (a) the order of matrix (b) the number of decimal digits desired in the results.

In preparing the problem tape, the correlations are punched by rows on a teletype key-punch. Each correlation is preceded by the sign (indicated by the sexadecimal digits corresponding to decimal '10' and '11', respectively) and by two scaling zeros (since no number greater than ±1.00 can be stored by the machine). The punched tape may be checked in either of two ways: (a) repunching the tape, and comparing the two versions on a teletype comparing unit; (b) printing the punched tape, and reading the printed version to some second person. In the event of any mistake, this is marked on the teletype tape. The tape is then fed through a copying unit which is stopped when the error is reached, enabling any insertion, deletion, or replacement to be made.

The electronic computer time for any task falls into three sections:

1) *Input.*—The teletype tape is read by the photo-electric reader. The principal axes tape for a matrix of order 23 is read in about twenty seconds.

2) *Calculation.*—With a matrix of order 23, more than two

thousand successive sine-cosine transformations will be made in diagonalizing the matrix. These computations are carried out in about seven minutes.

3) *Output.*—Results are punched on tape. The latent root for each factor is punched first, followed by the n factor loadings. After a space in the tape, the next factor is punched. About four minutes is required for this punching.

A printed copy of results is obtained by feeding punched results through the teletype unit at any convenient time. This prints three hundred sixty characters per minute (each sign and decimal digit is a separate character). Between fifteen to twenty minutes is needed for this, but the time may be reduced by ordering the computer to punch fewer than eleven decimals for each loading. Further copies may be printed if needed.

Machine errors are rare on an electronic computer. So far as our experience goes, they are much less frequent than in the use of punched-card equipment. A machine may run for a good many hours without making any mistakes; in that time an almost incalculable number of individual operations and separate factor analyses will be carried out. In the event of error, an entire analysis can probably be repeated in less time than would be needed for finding and correcting one single punched-card mistake.

Although errors are few, routine checking of results is, of course, essential. Checking may be carried out in either of two ways:

1) A desk calculator may be used to check that the norm and the trace of the correlation matrix have remained invariant under the process of diagonalization.

2) The electronic computer may be used to check that the correlations may be restored by the multiplication of the principal axes factor loadings. The tapes with correlations and factor loadings are fed into the machine, which is ordered to indicate whether the check agrees within specified limits of tolerance. A program for the Illiac to make this check is to be prepared.

The principal difficulty in the use of an electronic computer is in the avoidance of human error. Any human error probably requires the repetition of an entire analysis, whereas any error of the same sort on punched-card equipment or desk calculator may as a rule be rapidly amended. A further difficulty is that the handling of teletype tapes presents greater problems than the handling of IBM cards when there are large sets of data to be analyzed. A strong case can be made for the punching of the test measurements on to IBM cards rather than teletype tape, with the use of the IBM Type 63 machine for transfer of data from cards to tape. Errors are more easily corrected, and operators seem to attain greater speed on the IBM Key Punch. A case can sometimes be made for the transfer of final results back to cards by means of the IBM Type 43 machines, because of the superiority of IBM printing equipment over the teletype counterpart.

The electronic computer work makes possible the use of 'integrated programs'. Rather than calculate product-moment correlations in one machine operation and factor loadings in another, it is better for both to be calculated in a single machine run, once the two sets of orders have been integrated into one program.

SUMMARY

In view of the increasing use of electronic computers in psychological research and the resulting practicality of principal axes solutions in factor analysis, the procedure is given for the execution of a principal axes analysis on a computer. After discussion of the reasons for preferring a principal axes to a centroid solution in the initial factoring of the correlation matrix, some of the main features of the Illiac (the University of Illinois electronic computer) are described. There follows an account of the sine-cosine method which has been adopted for calculating the factor loadings, and of the day-to-day computational procedure for factorizing a correlation matrix once the appropriate set of machine orders has been prepared.

REFERENCES

1) A. C. Aitken. "Studies in practical mathematics. II. The evaluation of the latent roots and latent vectors of a matrix." *Proc. Roy. Soc. Edin.*, 1937, 57, 269–304.

2) M. S. Bartlett. "Tests of significance in factor analysis." *Brit. J. Psychol. Statist. Sect.*, 1950, 3, 77–85.

3) ——"A further note on tests of significance in factor analysis." *Brit. J. Psychol. Statist. Sect.*, 1951, 4, 1–2.

4) C. Burt. "The unit hierarchy and its properties." *Psychometrika*, 1938, 3, 151–168.

5) ——"Alternative methods of factor analysis and their relations to Pearson's method of 'Principal axes'." *Brit. J. Psychol. Statist. Sect.*, 1949, 2, 98–121.

6) F. B. Davis. "Fundamental factors of comprehension in reading." *Psychometrika*, 1944, 9, 185–197.

7) F. B. Davis. (Ed.) *The AAF Qualifying Examination*. AAF Aviation Psychology Program Research Reports, Report No. 6. Washington: D. C.: U. S. Government Printing Office, 1947.

8) K. J. Holzinger and H. H. Harman. *Factor Analysis*. Chicago: University of Chicago Press, 1941.

9) H. Hotelling. "Analysis of a complex of statistical variables into principal components." *J. educ. Psychol.*, 1933, 24, 417–441, 498–520.

10) ——"Simplified calculation of principal components." *Psychometrika*, 1936, 1, 27–35.

11) C. G. J. Jacobi. "Über ein leichtes Verfahren die in der Theorie der Säcularstörungen vorkommenden Gleichungen numerisch aufzulösen." *J. reine angew. Math.*, 1846, 30, 51–94.

12) T. L. Kelley. *Essential Traits of Mental Life*. Cambridge, Mass.: Harvard University Press, 1935.

13) C. Lanczos. "An iteration method for the solution of eigenvalue problem of linear differential and integral operators." *J. Res. nat. Bur. Stand.*, 1950, 45, 255–282.

STUDENT, SUPERVISOR, AND SELF-RATINGS OF INSTRUCTIONAL PROFICIENCY[1]

WILSE B. WEBB

Aviation Psychology Laboratory, U. S. Naval School of Aviation Medicine

and

C. Y. NOLAN

Human Factors Operational Research Laboratory

In the evaluation of teaching proficiency there are at least two classical methods of evaluation: student ratings and supervisor ratings. The present study is concerned with the relations between these evaluative procedures and a third evaluative source—the teacher's self-evaluation of his own proficiency.

Many arguments have been put forth in favor of and against student ratings and supervisor (or peer) ratings: "the student is the consumer and hence must be the judge;" "the student is incapable of evaluating what he is learning at the time;" "the supervisor is the only one capable of evaluating, from a broad base of experience and understanding, what is to be taught;" "the supervisor does not have the opportunity to observe or his observations are distorted;" etc. These arguments have been too frequent to review here and too frequently merely argumentative. We would like to comment on the self-rating as a source of evaluation because of its more infrequent use. Basically we feel that personal learning and improvement stems from an understanding of one's own adequacies and inadequacies. We feel that a self-evaluation procedure serves to focus the individual's attention on his inadequacies and as a result he will be motivated to attempt to correct them. Admissibly this technique cannot serve as an administrative device since the man will distort these ratings for secondary purposes. However, in a non-threatening situation and in conjunction with other evaluative procedures, we feel that self-ratings are a valuable adjunct to improving and evaluating teaching proficiency.

---

[1] The present research was performed as a part of ONR Contract 816(02). The opinions or conclusions contained in this report are those of the authors. They are not to be construed as necessarily reflecting the view or endorsement of the Navy Department.

PROCEDURE

A rating scale covering seven characteristics of importance to the teaching situation was administered to fifty-one instructors. The characteristics rated were interest of the instructor in his subject, his sympathetic attitude toward students, presentation of subject matter, sense of proportion and humor, self-reliance and confidence, personal peculiarities, and personal appearance. Each of these characteristics was rated on a zero- to ten-point scale.[2]

The procedure for administering the ratings follows. The problem was explained at length to small groups of instructors, and it was emphasized that none of the results of the ratings for any individual would be disclosed. The instructors were then asked to rate themselves. After additional instruction about avoiding biasing the students, the proper number of questionnaires were given each instructor. Since the students were not required to sign their names to their rating scale, the instructor was also given an envelope which was to be sealed after being filled with the completed student ratings. The instructor's name was then put on the envelope, and it was turned into the proper phase office. In all cases the students had at least eight hours of instruction before rating their teachers. The subject matter of the course was mathematics, physics, layout and hand tools.

A meeting of all the supervisors of the instructors used in the study was next called. The nature of the problem was again explained at great length and any questions answered. The supervisors were then asked to rate their instructors on the same form. In addition they were asked to rate the instructors on an Air Force forced-choice rating form, the Instructors Description Form C which was developed by Highland and Berkshire (*1*).

Biographical information about instructors was collected from various sources. These data included the GCT score of each instructor, the amount of his formal education, the number of years he had taught, and his enthusiasm for teaching as indicated by him on a seven-point scale. A number of product-moment

---

[2] This study was performed in the Naval Air Technical Training School at Jacksonville, Fla.

TABLE 1

Coefficients of correlation between:	
Instructor self-rating and student rating	.62
Instructor self-rating and supervisor rating	.16
Student rating and supervisor rating	.13
Instructor GCT score and instructor self-rating	−.25
Instructor GCT score and student rating	.21
Instructor GCT score and supervisor rating	.08
Instructor formal education and his own self-rating	−.23
Instructor formal education and student rating	.18
Instructor formal education and supervisor rating	.08
Instructor's teaching experience and his own self-rating	.05
Instructor's teaching experience and student rating	−.21
Instructor's teaching experience and supervisor rating	−.13
Instructor's enthusiasm for teaching and his own self-rating	.25
Instructor's enthusiasm for teaching and student rating	.39
Instructor's enthusiasm for teaching and supervisor rating	.00
Supervisor rating on Air Force form and on our form	.37
Supervisor rating on Air Force form and instructor self-rating	.20
Supervisor rating on Air Force form and student rating	.09
Student ratings in stanine form and difference scores between self-ratings in stanine form and student ratings in stanine form	−.13

NOTE: With a sample population of fifty-one the five and the one per cent levels of significance are 0.27 and 0.35, respectively.

correlations were computed between these data and those derived from the two rating scales. These coefficients are listed in Table 1.

DISCUSSION

Admittedly, the present study was not made in the typical educational setting. The teachers were not, for the most part, professional teachers. The supervisors were not principals or even trained professional supervisors. However, we feel our study contains certain generalities which should be considered by professional educators. After all, this was a classroom situation in which somewhat classical subject matter was being presented to typical students. Further, the supervisors were responsible for evaluation of instruction and had previously, formally and relatively continuously been rendering such judgments.

Our main findings seem to be rather clear-cut. There is a rather high relationship between the way the student views a teacher as a teacher and the way the teacher views himself. When the limits of reliability of group-ratings and self-ratings are taken into account, this relationship is strikingly high (*2*). In other words, the teacher does have an idea of himself which is quite similar to the 'consumer's' idea when he is called upon to make such an evaluation. We feel that this insight can be used as a prime source of instructional improvement.

When we turn to supervisor ratings we find a different picture. There is little relationship between the student's view of the teacher or the teacher's view of himself and the supervisor's ratings. In fact, it is difficult to tell what is the basis of the supervisor's ratings since these were not significantly correlated with the intelligence of the instructor (AGCT), his experience in teaching, his level of schooling, or his enthusiasm for teaching. We can only conclude that the supervisor was rating on some factor or factors other than those which were valid estimates of the teaching ability of the individual or were random invalid intuitions.

Certain interesting points can be noted about the student ratings and the instructor ratings. Only one further correlation obtained was statistically significant. This was the positive correlation of .39 between the instructor's enthusiasm for teaching and the student's ratings of his teaching ability. This would clearly support the hoary but apparently sound generalization that one of the prime attributes of a good teacher is his desire to teach. We may further note that, although not statistically significant, the next two highest correlations indicated the more intelligent and the more educated instructors seem to be more self-critical (correlations of −.25 and −.23 between the teacher's self ratings and the AGCT and level of schooling, respectively).

In the particular situation it would seem possible to state that the AGCT, the level of schooling or teaching experience (within the limits of the selected population) were not significant variables in the teaching situation.

The last correlation report in Table 1 was our greatest disappointment. We hypothesized that the greater the difference between the student rating and the self-rating, the lower the student rating would be. This was based on the assumption that widely disparate (assigned or judged) rôles between the student and the instructor would result in 'psychological friction'. As indicated by the −.13 correlation between these variables our hypothesis was not confirmed.

SUMMARY

Fifty-one instructors in a military setting were rated by their students and by their supervisors on a 'teaching proficiency' rating scale. In addition, the instructors rated themselves on the same scale. It was found that the student ratings and the self-ratings of the instructors were highly correlated. However, the supervisor's ratings were uncorrelated with any of the additional measures obtained (the instructor's AGCT, his level of schooling, his teaching experience, or his desire to teach). There was a tendency for the more intelligent instructors and those with more schooling to be more self-critical. The instructors who expressed a greater desire to teach were rated as superior teachers by their students. The discrepancy between student ratings and the instructor's ratings did not seem to be related to the judged proficiency of the teacher.

REFERENCES

1) R. W. Highland and J. R. Berkshire. *A Methodological Study of Forced-choice Performance Rating*. Air Training Command, Human Resources Research Center, Bulletin 51-9.

2) W. B. Webb and Billy Stafford. *Group Evaluations and Self-evaluations: The Matter of Reliability*. Unpublished study available from senior author.

RETROACTIVE INHIBITION IN MEANINGFUL MATERIAL[1]

JOHN F. HALL

The Pennsylvania State University

The experimental work in the area of learning has been frequently criticized from the point of view that there is little similarity between the materials used in such experiments and that found in the learning situation of the typical classroom.

This criticism is pertinent in the area of forgetting. Although most psychologists are in accord with the experimental evidence which indicates that retroactive inhibition makes a significant contribution to forgetting, most of the evidence demonstrating its operation has been done with either the traditional nonsense syllable, or the isolated noun or adjective. Those studies which have attempted to measure the susceptibility of meaningful material to retroactive inhibition have been few.

In one of the more carefully controlled studies in the area, McGeogh and McKinney (*3*) measured the susceptibility of meaningful material to inhibition generated from the interpolated learning of other meaningful material, as well as from nonsense syllables. The meaningful material was obtained from an elementary text in psychology. The results indicated that the differences in the amount recalled after varying interpolated materials had been presented (rest, nonsense syllables, and meaningful material) were not satisfactorily reliable. The failure to clearly demonstrate operation of the phenomenon would seem to indicate the need for additional research in the area.

The present study was a similar attempt to investigate the operation of retroactive inhibition with meaningful material employing two degrees of similarity between the original and interpolated material.

EXPERIMENTAL PROCEDURE

Subjects—A total of forty-five subjects, twenty-five male and twenty female college students, obtained from an introductory

---

[1] This study was supported by funds obtained from The American Philosophical Society.

course in Psychology, were used in the experiment. Fourteen students comprised the control group, fifteen were used in the first experimental group, and sixteen in the second.

Procedure and Materials—The conventional paradigm used in retroactive inhibition experiments was used, and was as follows: Control group: Original learning—rest—test for retention. Exp. group: Original learning—interpolated activity—test for retention.

The experimental sessions for each group were held in the evening in one of the classrooms of The Pennsylvania State University. No information was given the subjects other than telling them that they were going to participate in a learning experiment and advising them to do their best.

The same procedure was followed for all groups in presenting the original material. The prose used for this learning situation consisted of an introductory statement and thirty sentences written specifically for the experiment. Each sentence consisted of from five to eight words. The content of the material was obtained from anthropological sources and described the characteristics and customs of a hypothetical primitive tribe of people.[2] This type of material was used to reduce the possibility of subjects having familiarity with the material.

The material was presented by means of a slide projector with each slide containing a single sentence and being presented for five seconds. Approximately three minutes were required to present the total material. After the last slide was presented, the amount learned was tested by utilizing a completion test. Students were given five minutes to write in a key word, usually an adjective or noun, which had been omitted from each of the thirty sentences. Copying was eliminated by careful proctoring and by having the subjects seated in alternate seats.

The interpolated material was then presented. For the control group, this consisted of administering the Minnesota Spatial Relations test which consumed twenty-five minutes, and the DuBois-Bunch learning test (*1*) which took twenty minutes. With this latter test, it was necessary for the subject to recognize and record which of four key figures appeared twice in a pattern. These two tasks are far from a rest condition, but the kind of activity required to complete them seemed to be considerably remote from the kind of activity required in the original learning. The use of such activity was, of course, to prevent the subjects from rehearsing any of the original material which had been presented.

---

[2] The following sentences are representative of the material used: "The tribe is located in *northern* Africa." "Women perform most of the *domestic* tasks." (Italicized words represent the words which were omitted in the completion test.)

The interpolated prose material for Experimental Group 1 consisted of material describing the characteristics and customs of a second hypothetical primitive tribe. Thirty sentences were used to describe this material, with the length of each sentence being similar to sentence length of the original material. Every effort was made, however, to keep to a minimum the similarity of content between this material and the original.

The interpolated prose material for Experimental Group 2 consisted also of thirty sentences. Twenty sentences were identical to those used with Experimental Group 1. The other ten, however, were identical to ten sentences contained in the original material, except for the presence of a single word, and it was this word which was omitted in the completion test.[3] It was felt that these ten sentences of interpolated material, because of their similarity to the original material, would produce substantial amounts of inhibition of the original material. These ten items were distributed at equal intervals throughout the list of thirty.

The time required to present the interpolated prose material and to administer the completion test was approximately nine minutes.

In order to keep the amount of time spent in interpolated activities constant for all groups, the two experimental groups were administered also the Minnesota Spatial Relations test, and the DuBois-Bunch learning test. The time required to administer the Minnesota Spatial Relations test was twenty-five minutes; the time allowed for the DuBois-Bunch learning test was eleven minutes. Thus, for all groups interpolated activities required forty-five minutes.

---

[3] The following represents this relationship: (Exp. Group 1) Original material—"The tribe is located in *northern* Africa." Interpolated material—"The tribe is organized into *numerous* divisions.

(Exp. Group 2) Interpolated material: "The tribe is located in *southern* Africa."

(Italicized words represent words which were omitted in the completion test.)

After the interpolated material had been presented, the completion test covering the original material was presented. Five minutes were allowed for its completion.

Twenty-one days later, a completion test covering the original material was again presented, and five minutes permitted for its completion.

RESULTS AND DISCUSSION

Table 1 presents the results for each of the groups for the three testing periods. The results are in terms of (1) the mean number of words correctly recalled, thus completing the sentence, and (2) per cent of recall. This latter measure was obtained by dividing the mean raw score at immediate recall into the corresponding score at each delayed recall.

Although the findings indicate that Experimental Group 2 had forgotten the largest number of words, the number of omissions is not significantly larger than those made by the other two groups.

It was possible that the inhibition generated by the ten similar

TABLE 1.—NUMBER OF COMPLETION ITEMS CORRECTLY RECALLED

Group	Kind of Interpolated Material	N	Immediate		After 45 min.			After 21 Days		
			Mean	σ	Mean	σ	Per cent recalled	Mean	σ	Per cent recalled
Control	Rest	14	23.36	3.80	23.14	3.56	99.1	18.93	3.08	81.0
Exp. Gp. 1	Dissimilar material	15	23.20	3.27	23.53	3.34	101.4	19.33	3.93	83.3
Exp. Gp. 2	Similar material	16	23.31	2.14	22.94	3.84	98.4	18.81	4.59	80.7

TABLE 2.—NUMBER OF TEN COMPLETION ITEMS CORRECTLY RECALLED

Group	Kind of Interpolated Material	N	Immediate		After 45 Min.			After 21 Days		
			Mean	σ	Mean	σ	Per cent recalled	Mean	σ	Per cent recalled
Control	Rest	14	8.64	1.23	8.57	1.40	99.2	7.21	1.47	83.4
Exp. Gp. 1	Dissimilar items	15	9.07	1.24	9.07	1.39	100.0	7.47	1.36	82.4
Exp. Gp. 2	Similar items	16	8.63	1.05	8.63	1.27	100.0	6.69	1.79	77.5

items found in the interpolated material for Experimental Group 2 was not sufficiently large to produce significant differences between the groups in the recall of the original material. Consequently, a separate analysis of these ten items was made, and is indicated in Table 2.

Here again, significant differences in omissions do not appear at either the forty-five minute or the twenty-one day interval, although this latter recall period reveals Experiment Group 2 again had the largest amount of decrement.

It is interesting to note that the present results closely parallel those of McGeogh and McKinney (*3*). Not only do we find small amounts of inhibition generated by interpolated material, but find, also, almost equal potency of the three kinds of interpolated material on the inhibition of the original material. This is particularly puzzling inasmuch as we have deliberately introduced similar interpolated material, which, according to McGeogh (*2*), and the recent theorizing of Osgood (*4*), should have produced substantial amounts of inhibition of the original material.

It is apparent from these results as well as the results of others that more investigation is needed in this area in order to substantiate the hypothesis that retroactive inhibition makes an important contribution to the forgetting of meaningful material.

SUMMARY

This experiment was an attempt to demonstrate the operation of retroactive inhibition with meaningful material. The conventional RI paradigm was employed. Three groups of subjects were presented with meaningful material, and a completion test was utilized to measure the amount learned. Interpolated material was then presented, with Experimental Group 1 receiving dissimilar meaningful material, and Experimental Group 2 receiving similar meaningful material. A control group did not receive any meaningful interpolated material. Forty-five minutes and twenty-one days later, the amount of original material recalled was measured for all groups. Results indicated no significant differences among the groups.

REFERENCES

1) P. H. Dubois and M. E. Bunch. "A new technique for studying group-learning." *Amer. J. Psychol.*, 1949, 62, 272–278.

2) J. A. McGeogh and A. L. Irion. *The Psychology of Human Learning.* New York: Longmans, Green, 1952.

3) J. A. McGeogh, and F. McKinney. "The susceptibility of prose to retroactive inhibition." *Amer. J. Psychol.*, 1934, 46, 429–436.

4) C. E. Osgood. *Method and Theory in Experimental Psychology.* New York: Oxford Univ. Press, 1953.

NOTE ON THE VALIDITY OF THE ARMY GENERAL CLASSIFICATION TEST AS A PREDICTOR OF ACADEMIC ACHIEVEMENT

TOLAN L. CHAPPELL

University of Missouri

The lack of validity data and college norms for the Army General Classification Test limits its use by college personnel workers and administrators for counseling and selection purposes. One study (*1*) has appeared in the literature in which means, standard deviations, and validity coefficients are reported for various groups of officer candidates and ASTP cadets, but nothing has been reported with respect to use of the AGCT with college students in a civilian setting.

In the fall of 1951 the AGCT was administered to all University of Missouri male freshmen registered for the basic Army and Air Force ROTC courses. These men were enrolled in one of the four colleges which are regularly open to freshmen (Agriculture, Arts & Science, Education, and Engineering). At the end of the school year AGCT scores (Army standard scores) and first-year grade point averages were correlated. It was found that, for this group, the AGCT is about as good a predictor of academic achievement as any of the other tests used in the freshman placement test battery at the University of Missouri. This battery consists of: (1) The ACE Psychological Examination, 1942 edition; (2) Coöperative English Test A: Mechanics of Expression, Form S; and (3) Coöperative English Test B2: Effectiveness of Expression, Form S.

The test used in the present study was AGCT-1c which has been calibrated to AGCT-1a; the correlation between the two forms being .90 (*2*). AGCT-1a is the form which has been released for civilian use (*3*), and is now referred to as the first civilian edition, forms AH(hand scored) and AM(machine scored).

Validity data and college norms for AGCT-1c are presented in Tables 1 and 2, respectively.

TABLE 1.—VALIDITY COEFFICIENTS OF THE AGCT AS A PREDICTOR OF ACADEMIC ACHIEVEMENT

Group	N	Mean	SD	r
Agriculture	277	118.1	11.0	.47
Arts & Science	389	121.6	10.1	.43
Education	38	118.8	11.4	.39
Engineering	107	125.9	10.5	.36
Total	811	120.8	10.8	.41

TABLE 2.—PERCENTILE EQUIVALENTS FOR ARMY STANDARD SCORES ON THE ARMY GENERAL CLASSIFICATION TEST*

AGCT Score	Percentile	AGCT Score	Percentile
151–154	100	121	55
145–150	99	120	51
143–144	98	119	46
141–142	97	118	42
140	96	117	38
139	95	116	35
138	94	115	32
137	93	114	27
136	92	113	23
135	91	112	20
134	90	111	17
133	88	110	14
132	86	109	12
131	84	108	10
130	82	107	8
129	80	106	7
128	77	104–105	6
127	73	102–103	5
126	69	100–101	4
125	66	99	3
124	63	95–98	2
123	60	88–94	1
122	58	76–87	0

* Based on 992 University of Missouri male freshmen (1951–52).

REFERENCES

1) Staff, Personnel Research Section, A.G.O. "The Army General Classification Test." *Psychol. Bull.*, 1945, 42, 760–768.

2) Staff, Personnel Research Section, A.G.O. "The Army General Classification Test, with special reference to the construction and standardization of forms 1a and 1b." *J. educ. Psychol.*, 1947, 38, 385–420.

3) D. E. Super. *Appraising Vocational Fitness by Means of Psychological Tests*. New York: Harper and Brothers, 1949.

BOOK REVIEWS

Glenn Myers Blair, R. Stewart Jones and Ray H. Simpson, *Educational Psychology*. New York: The Macmillan Company, 1954, pp. 601. $4.75.

L. J. Cronbach, *Educational Psychology*. New York: Harcourt, Brace & Co., 1954, pp. 628. $5.50.

W. D. Commins and Barry Fagin, *Principles of Educational Psychology*. New York: The Ronald Press Co., Second Edition, 1954, pp. 795. $5.75.

H. H. Remmers, Einar R. Ryden and Clellen L. Morgan, *Introduction to Educational Psychology*. New York: Harper and Brothers, 1954, pp. 435. $4.00.

Henry P. Smith, *Psychology and Teaching*. New York: Prentice-Hall, Inc., 1954, pp. 466. $4.95.

Herbert Sorenson, *Psychology and Education*. 3rd Edition. New York: McGraw-Hill Book Company, Inc. 1954, pp. 577. $5.50.

It is a matter of no small interest that six textbooks in Educational Psychology should have appeared in the Spring of 1954. Two of these are editions of books that have been on the market for some years, but four are new. There is no way of knowing why there should be such an avalanche of texts in this field appearing within a few months of each other, but undoubtedly part of the explanation can be found in the fact that these books are good sellers and money-makers. A recent report on "Psychologists in Teacher-training Institutions"[1] shows that educational psychology texts are a close second to general texts in psychology used in these institutions. Undoubtedly the authors of these new texts have hoped to catch some of this market.

In order to determine the trends which these new books represent, a tabulation was made of the per cent of page space devoted to various topics to be compared with a similar tabulation of the page space of six books having wide use and popularity at the present time. These tabulations are reported in Tables I and II.

These six new books in Educational Psychology show marked differences from similar text available two and three decades ago.

---

[1] P. M. Symonds, S. Z. Klausner, J. E. Horrocks and V. H. Noll. "Psychologists in teacher-training institutions," *The American Psychologist*, 7: 24–30, 1952.

TABLE I.—PERCENTAGE OF PAGES DEVOTED TO TOPICS IN EDUCATIONAL PSYCHOLOGY, NEW 1954 BOOKS

Topic	Textbooks						
	Blair	Commins	Cronbach	Remmers	Smith	Sorenson	Average
Learning	18.4	28.9	25.1	10.2	21.1	30.5	22.3
Development (physical, emotional, mental)	12.2	7.1	3.0	6.8	14.4	15.8	9.9
Social development	7.7	14.9	5.0	15.5	7.2	5.3	9.3
Individual differences, intelligence	6.5	20.8	5.4	6.3	13.6	8.6	10.2
Motivation	7.5	4.5	10.0	0.8	5.3	7.9	6.[illegible]
Adjustment, personality, mental hygiene	14.5	9.1	12.8	30.4	15.9	13.2	16.0
Measurement and evaluation	19.5	5.4	14.3	30.2	6.2	12.1	14.6
Attitudes and interests	3.6		8.2		7.9		3.3
Psychology of the teacher	10.1		8.9		8.4	6.4	5.6
General orientation to education and psychology		9.3	7.3				2.8

There is practically no reference to the nervous system; indeed, most textbook writers have swung too far in the direction of proclaiming their ignorance of the exact connection between mental activity and nervous activity. At least, these reviewers hold that students should know that most psychologists believe that there is such a connection, even though the details have not yet been worked out. There is also little emphasis on the history of psychology, and most of these new textbooks give little mention to some of the early names such as G. Stanley Hall, William James, E. L. Thorndike, William MacDougall, or even more contemporary names, such as Clark Hull, C. E. Skinner and Edward Tolman. There is a tendency to minimize the several schools of psychology, and these texts pay little attention to theoretical issues which have interested experimentalists over the last decade. Most of these texts give a straight-forward explanation and description of learning without introducing technical terms or going into matters of theory and controversy.

From the tabulation we learn that the current crop of texts in Educational Psychology gives less space to topics of learning and development than those that preceded them. On the other hand,

TABLE II.—PERCENTAGE OF PAGES DEVOTED TO MAIN TOPICS IN EDUCATIONAL PSYCHOLOGY, BOOKS IN CURRENT USE

Topic	Textbooks*						
	Cole	Gates	Pressey	Skinner	Sorenson	Trow	Average
Learning	31.7	15.6	32.6	29.8	30.3	41.7	30.2
Development (physical, emotional, mental)	16.4	21.2	19.2	15.0	16.3	12.2	16.7
Social development	5.3	6.8	11.0	4.0	4.0	15.4	7.7
Individual differences, intelligence	7.9	5.1	8.2	12.3	16.7	6.3	9.4
Motivation	6.2	1.5		3.0	9.4	6.3	4.4
Adjustment, personality, mental hygiene	13.8	29.3	7.4	21.4	10.6	5.6	14.7
Measurement and evaluation	8.8	13.2	5.6	7.3	9.5	5.6	8.3
Attitudes and interests			16.1	3.6			3.3
Psychology of the teacher	4.0	5.1			3.2	4.3	2.8
General orientation to education and psychology	5.7	1.9		3.4		2.2	2.2

* L. E. Cole and W. F. Bruce, *Educational Psychology*. Yonkers-on-Hudson, World Book Co., 1950. A. I. Gates, A. T. Jersild, T. R. McConnell and R. C. Challman, *Educational Psychology*, 3rd Edition, New York: The Macmillan Co., 1923, 1930, 1942, 1948. S. L. Pressey and F. P. Robinson, *Psychology and the New Education*. New York: Harper and Bros., 1933, 1944. C. E. Skinner, Editor, *Educational Psychology*, 3rd edition. New York: Prentice-Hall, Inc., 1936, 1945, 1951. Herbert Sorenson, *Psychology in Education*, 2nd edition. New York: McGraw-Hill Book Co. Inc., 1940, 1948. W. C. Trow, *Educational Psychology*, 2nd edition. Boston: Houghton Mifflin Co., 1931, 1950.

space is distributed over a somewhat wider range of topics, which means that slightly more emphasis is given to problems of measurement and evaluation, motivation, adjustment, personality, mental hygiene, social psychology, and the psychology of the teacher. One very important trend that these reviewers have noted is the attempt to breach the gap between theory and educational practice. This has perhaps been one of the most cogent criticisms of educational psychology texts in the past—that they have reported well the results of psychological experiments, but that they had not translated these results into suggestions that would be helpful to the teacher. In these texts there is less reporting of the results of the experimental studies and more interpretations of the results of these studies in terms of educational practice and procedure.

The book by Blair, Jones and Simpson represents a rare piece of teamwork. We are told in the Preface that Blair has been responsible for the introductory chapter, Part II dealing with growth and development and three of the four chapters in Part IV, dealing with adjustment and guidance. Jones has taken as his responsibility Part III, dealing with learning as well as Chapter 16 on studying the individual child. Simpson has been responsible for Part V on measurement and evaluation and Part VI on the psychology of the teacher. This distribution of responsibility in the book is somewhat puzzling to these reviewers because each author seems to have gone out of character somewhat, if one may judge by his previous writings and interests. For instance, Blair is best known through his earlier book on *Diagnostic and Remedial Teaching in Secondary Schools*,[2] so that one would have expected him to have taken responsibility for chapters on learning and applications of learning to classroom situations. Likewise Simpson's earlier book, *Improving Teaching-Learning Processes*,[3] would lead one to expect him also to have contributed chapters on the application of the principles of learning to teaching procedures instead of chapters on measurement and evaluation for which he is responsible in the present book. On the other hand, Jones' previous writings, including his doctoral dissertation dealing with *Integration of Instructional with Self-scoring Measuring Procedures*,[4] would have led one to expect him to have written chapters on measurement and evaluation instead of the chapters on learning.

The merit of this book resides in the interpretation of experimental results and theory into principles of teaching and applications to classroom procedures. These authors not only present the results of studies that have been reported in the literature, but tie these findings to applications to the school situation.

Cronbach's book was first tried out in mimeographed form in his eral sections of the book is not so clearly apparent. We are not told in the Preface the responsibility of each of the three authors, but there is evidence that the senior author has been influential throughout, if indeed he has not actually written chapters in most of the sections of the book. Morgan apparently has contributed from his experiences with the Veterans Administration to the first six chapters in the book dealing with adjustment. A seventh chapter, dealing with a sociological perspective on education, has been prepared by Schneider. The two chapters on learning, we are told, were written with the assistance of Floria Fischer.

---

[2] G. M. Blair. *Diagnostic and Remedial Teaching in Secondary Schools.* New York: The Macmillan Co., 1946.

[3] R. H. Simpson. *Improving Teaching-Learning* Processes. New York: Longmans, Green & Co., 1953.

[4] R. S. Jones. *Integration of Instructional with Self-scoring Measuring Procedures.* Unpublished Ph. D. dissertation, Ohio State University, Columbus, Ohio, 1950.

This book is not designed with the needs of teachers exclusively in mind, as there are chapters dealing with the extent to which home conditions and the community can contribute to adjustment.

The last part of the book is heavily weighted with an emphasis on individual differences, problems of testing, and the statistical treatment of test results. These reviewers are a little puzzled by the introduction of a chapter on "Interpreting Test Scores" in an elementary text in educational psychology, and wonder to what extent the authors expect the student in a beginning psychology class to master the elements of statistical methods. The inclusion of an appendix in the book, dealing with measures of relationship, seems even less appropriate. Indeed, the book is very heavily weighted with statistical concepts. For instance, on page 1 we are told that "psychologically a normal individual is at best a statistical fiction," and in the chapter on "How and Why We Learn," in the section dealing with reasoning, we find the following given as an illustration: "Let us now look at a student's effort to enlarge upon or formulate a concept. Let us take as an example the concept of an average as exemplified by the formula: average = $\Sigma m/N$ (sum of all the measures divided by the number of measures)." (p. 269).

Certain portions of this book lead the reviewers to the impression that it was written several years ago. For instance, in the chapter on "Achievement Testing" with one sole exception the last reference is dated 1941. Perhaps the authors feel that there has been little added to the literature on achievement testing in the last thirteen years. There are sections in this book that make one believe that the authors have not kept up with contemporary developments in psychological thinking. For instance, on pages 280 and 281 it is clearly implied that reward and punishment operate in opposite directions, whereas it was in 1929 that Thornclasses over a period of years and whatever merits the book possesses may be attributed in part to the suggestions and criticisms made by his students. Cronbach has done a superb job in bridging the gap between theory and practice, and has made notable steps forward in interpreting the results of experimental studies in terms of their classroom applications. These reviewers find the chapters on "Purposes and Aspirations," "The Teacher as a Classroom Leader," and "Personality Structure and Character Development" excellent: these chapters representing a somewhat new emphasis in texts of this kind. Since Cronbach is recognized as a leader in measurement, statistics and research, one might expect an overemphasis of these topics in the book. On the contrary, measurement has been fitted in without undue emphasis, problems of the measurement of ability and aptitude being taken up under the general heading of 'readiness', and problems of the measurement of achievement being taken up under the heading of 'evaluation'. The book itself is attractively published, and there is evidence that both author and publisher took a great deal of pains in the manufacture of the book. There is a copious use of pictures, but the reviewers feel that, contrary to the use of pictures in some current textbooks, Cronbach has made his pictures really instructive and they have added to the text itself. There are summaries, questions and reading lists at the end of the chapters which make the book of greater use to both the instructor and the student.

The book by Commins and Fagin represents a thorough revision and rewriting of their earlier book of seventeen years ago. This text is one of the most scholarly of the group. There are references to historical developments in psychology and attention is paid to the theoretical differences between the several schools of psychology. Careful definitions of terms are given.

The emphasis in this text is on gestalt and field theory. One notes that the heroes of these two writers are Tolman, Wheeler and Perkins, Humphrey, Hartmann, Lewin, and other field theorists. These reviewers found the chapters on thinking processes and problem-solving particularly valuable. It is of note that those who approach these topics from the point of view of connectionistic psychology have less to contribute than those who approach them from the point of view of field theory.

The book by Remmers, Ryden and Morgan also represents the team approach. However, in this book the integration of the sevdike clearly stated that punishment did not result in the diminution of strength of a connection in the same way that a reward helps to strengthen a connection.

These authors have provided an elaborate study manual to be used in connection with the text.

The book by Smith is described on the jacket as a new approach to educational psychology. Unfortunately with so many competitors this book cannot claim quite so much originality, but at the same time it has several features which make it desirable as an elementary text. Smith has avoided giving attention to the controversies that command the attention of advanced students. He has chosen to omit any historical orientation to the topics in educational psychology and he has written with a minimum of technical vocabulary. This is perhaps one of the most elementary of the texts being reviewed and this probably was the author's intention. There is a good balance in the distribution of topics in the book between development, learning, and motivation. An introductory chapter deals with the teacher as a professional worker, reporting some of the characteristics of the good teacher and some of the problems most frequently met by teachers, as an introduction to the main topics in the book itself. The topic of evaluation is limited to one chapter. This author has been as successful as any in interpreting the scientific principles which have been developed over the years in terms of classroom practice.

The book by Sorenson, which is the third revision of this widely used text, represents the author's and publisher's continued efforts to keep the text abreast of the times. This book has been collated with the McGraw-Hill Text-Films on Educational Psychology. The arrangement of topics is somewhat unusual. The book begins with a chapter on "Measuring the Learning and Achievements of Pupils." We are told that this chapter is placed early so that the instructor can give his attention at the beginning of the course to the basis on which he will assign marks to students in his course, and so that the students will know how they are to be graded. Chapter III deals with the problem of study, and this, too, has been placed early because the author believes that students need early guidance and instruction in the art of study. In general, the problem of growth is taken up in the book before learning. Physical growth is treated before emotional growth, and the problems of intelligence and its measurement and its relation to heredity precede the discussion of learning. Each chapter is introduced by a section entitled "What To Look For" that raises questions which the student is to keep in mind, the answers to which are to be found in the text. Following this section there is an introductory paragraph in each chapter giving an anecdote or incident which serves as a kind of text for the subsequent discussion. At the end of each chapter there is a summary and review and finally a brief "Test Your Thinking." At the end of the book there is a glossary and a very ample bibliography.

In comparing this third revision with the revision which preceded it, we find less attention given to individual differences in intelligence and more attention to problems of adjustment, personality and mental hygiene. Sorenson also gives greater space to measurement and evaluation in this last revision and also to the psychology of the teacher. Supplementary materials have been prepared including a teachers' manual which contains materials which can be used for testing and a workbook to be put into the hands of students to help them in outlining and applying materials presented in the text.

It will perhaps be superfluous for these reviewers to predict trends which they foresee in the next batch of new and revised educational psychologies, but here are some of the trends as we see them. In the first place, there must be an inclusion in our texts on educational psychology of some of the recent developments in social psychology, particularly those having to do with group dynamics, the psychology of the group process both in small groups and large groups. Some of the current findings of problems of leaderships, of group morale and of group thinking must find their way into our educational psychology texts. Greater attention will be paid in the future to problems of teacher-pupil relationships, and some of the findings coming out of psychotherapy dealing with problems of transference and resistance have their direct applications to the classroom situation. Indeed, there should be further applications of principles of motivation and dynamic psychology to the problems of teaching and learning. Finally, there are new developments in our understanding of the rôle and function of the teacher as a person and of the part that dynamic factors in the teacher's personality play in the teaching situation. These new emphases and topics must be not merely superimposed upon but integrated with the principles that have already been developed concerning development, learning, individual differences and the problems of evaluation.

PERCIVAL M. SYMONDS
ARTHUR R. JENSEN

Teachers College, Columbia University

THE JOURNAL OF

EDUCATIONAL PSYCHOLOGY

Volume 46 February, 1955 Number 2

THE EFFECT OF ORGANIZATIONAL AIDS ON COMPREHENSION AND RETENTION[1]

C. M. CHRISTENSEN

University of Arkansas

and

K. E. STORDAHL

Arkansas Polytechnic College

Many educators assume that an overview or preview of the main ideas in a printed passage will have a beneficial effect on comprehension. For example, one of the essential steps in the study technique developed by Robinson (7) includes a survey of the material prior to reading it. The same idea has also been expressed in terms of the 'whole-part-whole' method of learning. This method assumes that an individual will learn more efficiently if he groups the major ideas in order to get a clear picture of the structure of the material before beginning to learn the details. After mastering the details, the individual again considers the learning material as a whole.

Intuitively, it seems that an overview of learning material prior to its detailed consideration should aid comprehension. One might consider comprehension of textbook material as a continuous process of outlining or coding minor points under major points. Undoubtedly, people cannot retain great masses of information without some system for organizing it into coherent bodies or groups. A simple example of this is the retention of a telephone number such as MA 7168. It is much more difficult to remember this number if it is considered as six discrete elements, M-A-7-1-6-8, rather than three main elements such as Main-71-68 where each main element is composed of two subordinate elements.

---

[1] This experiment was made possible by Contract AF 18(600)-335 issued to the Division of Communications at the University of Illinois, and it was monitored by the Human Resources Research Institute. The authors wish to acknowledge the help of Dean Wilbur Schramm, Dr. C. E. Swanson, Mrs. Mary Snoden, and Mr. Herbert Rabin.

If this reasoning is valid, written materials with a clearly indicated organization should be more easily comprehended than materials without a clearly indicated organization. A number of methods have been used to indicate the organization of written materials. These methods have met with varying degrees of success.

There has been very little research directly related to this problem. The few studies that have appeared in the literature have in most cases been poorly designed so that clear cut conclusions cannot be drawn. McClusky (*5*), using college students as subjects, conducted an experiment to determine the effect of preliminary skimming on reading. The groups which did the preliminary skimming of headings and summaries tended to read faster than the control group. However, the controls and statistical analysis were inadequate.

Questions at the beginning, end, and within material have also been used to emphasize the main ideas. Holmes' (*2*) and Washburne's (*9*) studies suggest that the use of questions to guide the reader will facilitate comprehension. Unfortunately, clear cut conclusions cannot be drawn from these experiments. Robinson and Hall (*8*), in a study using materials with and without headings, found no differences in speed of reading and comprehension.

Dearborn, *et al* (*1*) suggest, on the basis of a study involving college students, that manuscripts be prepared in the following manner in order to facilitate reading speed and comprehension:

1) indicate the word in each sentence, which would be stressed when read aloud, by capitalizing it,

2) break sentences up into units of thought and

3) indicate important sections by overtyping on them. These suggestions emphasize the pattern or structure of the material.

The present study was concerned with different methods of indicating the organization or structure of printed material with reference to the differential effects of these methods on immediate and delayed retention. Outline, underlining, summary, and headings were the four methods used to indicate the organization of the material. This study was concerned with superimposing organizational cues on already written material. It did not consider training of individuals to organize material nor did it consider the rearrangement of the material itself.

METHOD

Experimental Materials.—The printed material used in this study consisted of two passages selected from United States Air Force Extension Course Institute volumes. One passage of approximately 2600 words described some of the elementary principles of aerodynamics. The other passage of approximately 3800 words considered the ideology of international communism. Two different kinds of passages were intentionally selected in order to provide a broader base from which to generalize.

Criterion Measures.—Two multiple-choice comprehension tests (Aero A, Aero B, Comm A, Comm B) were constructed for each experimental passage. Forms A and B were approximately equivalent for each passage. An effort was made to construct items which measured understanding of concepts and the ability to make inferences from general principles. The preliminary aerodynamics tests had thirty-six items and the communism tests had thirty-eight items.

The tests were pretested on a college freshman population and an item analysis was conducted using the upper and lower twenty-five per cent. The items which did not discriminate at the five per cent level were either eliminated or revised. Distractors which did not function were also revised and a few new items were added to make a total of forty items per test.

After the experiment was conducted, the difficulty of the items and the reliabilities of the tests were determined. The item difficulty ranged from twenty per cent to eighty per cent with a mean difficulty of approximately fifty per cent. The reliability of the tests is reported in Table 1.

Since there were only two forms of each test, Form A was used as both a pretest and delayed test. Consequently two reliability coefficients are reported for Form A of the tests.

Experimental Subjects.—The subjects used in this experiment consisted of nine flights of basic trainees at Sampson Air Force Base. All the subjects had been inducted into the Air Force less than two weeks prior to the experiment. The Air Force Qualifica-

TABLE 1.—SPLIT-HALF RELIABILITY (CORRECTED FOR ATTENUATION) OF THE COMPREHENSION TESTS

Tests	N	r (corrected)
Comm A (Pretest)	465	.62
Comm A (Delayed Test)	407	.70
Comm B (Immediate Test)	464	.76
Aero A (Pretest)	467	.61
Aero A (Delayed Test)	405	.75
Aero B (Immediate Test)	466	.85

ion Test (AFQT) percentile ranks for this sample yielded a mean of 55.3 and a standard deviation of 28.9. Number of years of education yielded a mean of 11.0 and a standard deviation of 1.7.

Experimental Versions.—The methods used to indicate the organization of materials were as follows: (a) an outline at the beginning of the passage, (b) a summary at the beginning of the passage, (c) a summary at the end of the passage, (d) underlining of main points, (e) headings in statement form, and (f) headings in question form.

The steps followed in constructing the experimental versions were:

1) The 'main points' were identified in each passage. A 'main point' was defined as follows: the definition of a word if understanding of the word was fundamental to understanding the passage, a statement of principles, a generalization, a proposition from which other things could be deduced, or a deduction or conclusion made by the author.

2) An outline, based upon the identified main points, was constructed. Detailed items such as examples, illustrations, and analogies were not included. The statements in the outline were made as brief as possible without distorting their meaning.

3) A summary was written. This summary contained approximately the same information as the outline.

4) The main points of each passage were underlined. An attempt was made to underline the same points as those included in the outline.

5) Headings were inserted in each of the passages. An attempt was made to keep the question and statement forms parallel and to base the headings on the outline. The headings indicated what

was to be discussed in the succeeding sections, but did not summarize the material. In this respect the headings differed from the other treatments.

The experiment was designed as a 3 × 3 × 2 × 2 factorial experiment. The design permits the comparison of all possible combinations of the organizational aids. This resulted in thirty-six different versions for each passage. A brief statement of each version is as follows:

1) Summary—beginning
2) Summary—end
3) Outline
4) Underlining
5) Headings—statements
6) Headings—questions

7) Headings—questions; summary—beginning
8) Headings—questions; summary—end
9) Headings—questions; outline
10) Headings—questions; underlining
11) Headings—statements; summary—beginning
12) Headings—statements; summary—end
13) Headings—statements; outline
14) Headings—statements; underlining
15) Summary—beginning; underlining
16) Summary—beginning; outline
17) Summary—end; underlining
18) Summary—end; outline
19) Outline; underlining

20) Headings—statements; summary—beginning; underlining
21) Headings—statements; summary—end; underlining
22) Headings—statements; summary—beginning; outline
23) Headings—statements; summary—end; outline
24) Headings—statements; outline; underlining
25) Headings—questions; summary—beginning; underlining
26) Headings—questions; summary—end; underlining
27) Headings—questions; summary—beginning; outline
28) Headings—questions; summary—end; outline
29) Headings—questions; outline; underlining
30) Summary—beginning; outline—underlining
31) Summary—end; outline; underlining

32) Headings—statements; summary—beginning; outline; underlining
33) Headings—statements; summary—end; outline; underlining
34) Headings—questions; summary-beginning; outline; underlining
35) Headings—questions; summary—end; outline; underlining

36) No organizational aids

of the material. In all instances, further analysis of the data did not seem warranted. All the simple analysis of variances were nonsignificant at the five per cent level and an inspection of the tables indicated that mean gains between the various treatments were extremely small.

The assumption of homogeneity of variance was tested for each analysis using Bartlett's test, Johnson (*3*, pp. 83–84). In no instance was the null hypothesis of equality of variances rejected at the one per cent level. The assumption of normality of distribution was checked by an inspection of the plotted data. The data satisfactorily conformed to this assumption.

In order to permit analysis of the results as a factorial experiment, a few subjects were eliminated for some of the analysis to give equal number of individuals per cell. This was done by a random procedure.

Immediate Retention.—The analysis of the differences between the experimental versions for the communism passage is presented in Tables 2 and 3. Table 2 contains the pre-test and immediate test scores for each of the experimental versions. Table 3 contains the results of the analysis of variance and covariance. It is clearly evident from an inspection of these tables that the differences between the experimental versions do not approach significance.

A similar analysis was performed on the results for the aerodynamics passage and, again, the differences were not significant at the five per cent level of confidence. Tables containing the data for this analysis and the remainder analyses considered have been omitted because similar procedures were employed.

TABLE 3.—ANALYSIS OF VARIANCE AND COVARIANCE OF THE COMMUNISM IMMEDIATE TEST SCORES WITH PRE-TEST SCORE HELD CONSTANT

Source of Variation	D.F.	S.S.		Cross Products	Adjusted Analysis				
		Pre-test	Immediate Test		D.F.	S.S.	M.S.	F	P
Between	35	1186.78	699.76	646.98	35	610.10	17.43	—	> .05
Within	396	13026.17	8550.90	6346.84	395	8315.28	21.05		
Total	431	14212.95	9250.66	6993.82	430	8925.38			

Bartlett's Test of Homogeneity of Variance: chi-square = 47.808; df = 35; P > [illegible]5.

Delayed Retention.—The statistical analysis for the delayed testing was identical to that for the immediate testing. No significant differences were found for either the communism or aerodynamics passages.

Time Devoted to Studying the Passages.—An analysis, similar to that described previously with the exception that analysis of covariance was not involved, was made to determine if subjects devoted more time to some experimental versions than others. This was a likely possibility as the reading material varied somewhat from version to version. For example, contrast the number of words in the version consisting of an outline and summary with versions having no outline or summary. No significant differences were found for either passage.

DISCUSSION

The consistency of the results, i.e., no significant differences, greatly simplifies any discussion. Clearly, one can say that the organizational aids used in this experiment did not effect comprehension as measured by the comprehension tests. Of course, one can only generalize to the population studied and the kinds of materials used.

Under different circumstances, organizational aids might very well effect comprehension. There are unlimited possibilities. A different population, e.g., college students, might respond favorably to organizational aids, or perhaps organizational aids are only useful for certain kinds of materials. Also training individuals to use the aids might result in a differential effect of the treatments used in this study. According to Robinson (*6*), many individuals do not study efficiently.

It is also possible that the experimental materials were not suitable for this particular population. The passages and tests may have been too difficult, although this seems unlikely in view of the item analysis and mean scores of approximately fifty per cent of the total possible score. In addition, the test items were carefully constructed.

SUMMARY

Two passages of material, one on international communism and the other on the principles of aerodynamics, were selected from Air Force correspondence courses. These passages were revised to include the following organizational aids: (1) an outline at the beginning of the passage, (2) a summary at the beginning of the passage, (3) a summary at the end of the passage, (4) underlining of main points, (5) headings in statement form, and (6) headings in question form. Thirty-six different experimental versions, using all combinations of the organizational aids, were used for each passage.

Two comprehension tests, Form A and B, were constructed for each passage. An item analysis was conducted and the split-half reliabilities determined.

A sample of Air Force basic trainees was selected as subjects. The subjects were first given a pre-test on the experimental materials; eight days later, they studied the passages and completed an immediate test of comprehension. Six days after the immediate testing, they were administered a delayed-test. The subjects also recorded the amount of time devoted to studying the passages.

Analysis of variance and covariance was applied to the results. The pretest scores were used in the analysis of covariance. No significant differences were found between the organizational aids for either immediate or delayed retention. No significant differences were found between time devoted to studying the different versions.

An explanation of these findings was difficult inasmuch as an unlimited number of explanations might be correct. Alternative explanations can only be determined by further research.

REFERENCES

1) Walter F. Dearborn, Philip W. Johnston, and L. Carmichael. "Improving the readability of typewritten manuscripts." *Proceedings of Nat. Acad. Sci.*, Wash., 1951, 37, 670–672.

2) E. Holmes. "Reading guided by questions versus careful reading without questions." *School Rev.*, 1931, 39, 361–371.

3) P. O. Johnson, *Statistical Methods in Research*. New York: Prentice-Hall, 1949.

4) M. G. Kendall, *The Advanced Theory of Statistics*. Vol. 2. London: Charles Griffin, 1948.

5) H. Y. McClusky. "An experiment on the influences of preliminary skimming on reading," *J. of Educ. Psych.*, 1934, 25, 521–529.

6) F. P. Robinson. "Study skills of soldiers in ASTP," *School and Soc.* 1943, 58, 398–399.

7) — — —. *Effective Study*. New York: Harper & Brothers, 1946.

8) F. P. Robinson, and P. Hall. "Studies of high-level reading abilities." *J. Educ. Psychol.*, 1941, 32, 241–252.

9) J. N. Washburne. "The use of questions in social science material." *J. Educ. Psychol.*, 1929, 20, 321–359.

A STUDY-ATTITUDES QUESTIONNAIRE FOR PREDICTING ACADEMIC SUCCESS

WILLIAM F. BROWN and WAYNE H. HOLTZMAN

The University of Texas

Publication of more than a hundred how-to-study manuals since 1926 attests to the strong interest of educators in the rôle of study habits and attitudes in determining academic success. However, the prolixity of rules found in these pamphlets and books has often lacked pertinent scientific support. Laycock and Russell (*7*), in analyzing thirty-eight how-to-study manuals published between 1926 and 1939, compiled a list of three hundred and thirteen items that were mentioned four or more times in the various manuals. Examination of the individual items revealed that there was direct disagreement in some areas as to the best techniques to use in studying. In an investigation of 1250 school children, Cuff (*3*) discovered that many of the seventy-five most commonly recommended principles of study were reportedly being followed more exactly by the inferior than by the superior students! Although such results may be due in part to inaccurate reporting by the students of their study habits, the urgent need for sound empirical investigations of the work methods adopted by good and poor students is readily apparent.

Several study-habits questionnaires have been published during the past twenty years which have been based upon analyses of the responses of good and poor students, but none thus far have shown any appreciable correlation with grades in school.

In 1933, Wrenn (*17*) published one of the first study-habits inventories. His final selection of statements was based upon an item analysis of the responses to a check list of study habits made by high and low scholarship groups matched for such important variables as intelligence and major field of study. In spite of the empirical basis for Wrenn's *Study-Habits Inventory*, subsequent validity studies by Reeder (*12*) and Gordon (*5*) indicate that only a slight relationship exists between scores on the inventory and scholastic success. A later item analysis of the *Study-Habits Inventory* by Wrenn and Humber (*18*), using the same methods as

Wrenn, prompted a revision of the original inventory, the validity of which has yet to be convincingly demonstrated.

A self-rating scale consisting of fifty-six statements was assembled by Locke (*8*) and administered to eight groups of college students with only fair success. Correlations between scores on Locke's *Student Skills Inventory* and one-semester grades ranged from −.07 to .39 with a median value of .28, only slightly higher than the correlations reported for Wrenn's inventory. Similar low but statistically significant relationships between scores on inventories of study habits and measures of scholastic success have been reported by Brooks and Heston (*1*), Schlesser and Young (*13*), and Michael and Reeder (*10*). Even less promising have been the studies by Eurich (*4*), Brown (*2*), and Schultz, et al (*11*, *14*).

For the most part, these investigations have aimed at uncovering the mechanics and skills of studying which might differentiate the good from the poor student. In general, these efforts have been unrewarding.

On the other hand, some of these studies have suggested that the degree of steady, highly motivated effort put forth by the student and his attitudes toward school work, rather than the purely mechanical procedures of studying, might be of considerable importance. In a three-year clinical study of girls at Sarah Lawrence College, Tiebout (*16*) concluded that the presence of a 'lazy student syndrome' was characteristic of students whose scholastic records were poorer than aptitude test scores would predict. This consisted mainly of four personality characteristics: (1) a need to rely upon strong and immediate motivations to start studying, (2) a tendency to have interests of a transitory nature, (3) a tendency to be governed by strong hedonistic principles, and (4) a deep-seated problem in learning. Other significant characteristics of the poor student as noted by Tiebout were a superficial and haphazardly disorganized quality pervading written work, a tendency to gloss over failures, to rationalize poor achievement, to procrastinate and yet show continual optimism about changing for the better.

Other, less systematic, investigations by Symonds (*15*) and Mathews and Toepfer (*9*) likewise suggest that the difference between good and poor students of comparable intelligence is mainly one of attitude, set, and motivation, rather than of professed study habits.

The present investigation was undertaken in an effort to construct and validate a self-rating questionnaire that would measure a student's attitude and motivation toward studying as well as his study habits *per se*. From group discussions with college freshmen, existing inventories on study habits, studies using observational and interview techniques to differentiate good and poor students, and reports on related experiments in the field of learning, two hundred and thirty-four statements were compiled. These statements were separated into two groups, one consisting of 'study-mechanics' items, the other of 'study-attitudes' items. This classification was based upon the authors' judgment as to whether an item elicited mainly degree of adherence to the mechanics and conditions of studying, or whether it was chiefly concerned with attitudes toward studying and motivation to do well in academic work.

Ambiguous statements were reworded or eliminated. The eighty-eight 'study-attitudes' and one-hundred 'study-mechanics' items thus obtained were assembled randomly in questionnaire form. Following each item a five-point scale was provided for the student to indicate whether he rarely, sometimes, frequently, generally, or almost always followed the procedure in question. The terms marking the intervals of the five-point scale were defined on a percentage basis to provide the subject with a definite standard for making judgments.

The preliminary questionnaire with accompanying IBM answer sheet was individually administered to one hundred and seventeen second-semester freshmen at The University of Texas in the Spring of 1951. All the subjects were carefully selected from either the dean's list of honor students or from the names of students on scholastic probation, representing good and poor students, respectively. To increase the likelihood of obtaining true responses, the purpose of the questionnaire was disguised by telling each student he had been selected at random from the freshman class to participate anonymously in an investigation of sex differences in study habits. In addition to standard instructions for filling out the questionnaire, each student was asked to circle the number on the answer sheet which corresponded to an item that he found difficult to answer. Any concern expressed by the student regarding the meaning of any word or phrase was recorded for future reference. In this way, information was obtained which later proved of value in refining the questionnaire.

Twenty-two of the twenty-three low-scholarship freshmen were then paired individually with twenty-two of the ninety-four high-scholarship freshmen. The members of each pair were matched as closely as possible on the following seven variables: sex; marital status; length of time in college; total score on the *American Council on Education Psychological Examination*, 1941 edition (ACE); date of birth; number of semester-hours of credit carried; and type of residence. Of the forty-four students, twelve were women and thirty-two were men.

Each item on every questionnaire was scored by assigning numerical values ranging from 1 to 5 to each of the five possible choices, a value of 1 being given to 'rarely', 2 to 'sometimes', and so on. Response-distributions for the high- and low-scholarship groups were compared for each item, using Fisher's *t* test to determine the significance of the difference between each pair of correlated means. Four of the one hundred and eighty-eight items differentiated the high and low groups beyond the .01 level of significance; nine proved significant at the .02 level; and thirteen were significant at the .05 level. Of special interest is the fact that of sixty items significant beyond the .20 level, forty-two were from the eighty-eight 'study-attitudes' items while only eighteen were from the one hundred 'study-mechanics' items.

This exploratory study suggested strongly that a self-rating questionnaire heavily loaded with statements referring to study-attitudes would be superior to the usual study-habits inventory for predicting the level of scholastic success in college.

Since it was desirable to reduce the average testing time from thirty-five to twenty minutes, the questionnaire was reduced in length to one hundred and two items. The rejection of items was based in part upon the exploratory item analysis, in part upon a tabulation of the number of times each item was marked as difficult to answer, and in part upon a subjective belief that some items contributed little to the questionnaire's purpose. In addition to eliminating eighty-six items, many of the retained items were reworded, and the sequence of the items was revised.

This 102-item questionnaire was administered in May, 1951, to two hundred and eighty-five men and two hundred and eighty-five women who were first-semester freshmen at San Antonio College, a municipal junior college; Trinity University, a church-related (Presbyterian) college; and The University of Texas. This was done at regularly scheduled meetings of freshman English classes in groups of twenty to thirty-five subjects each. All students in beginning English courses at San Antonio College and Trinity University and a random sample of similar classes at The University of Texas were tested to insure a representative sample of first-semester freshmen. Only unmarried students less than twenty-four years of age who were in their first semester of college work were included in the study.

The letter grades earned and semester hours of credit taken by each student during the Spring semester of 1951 were used as a basis for selecting good and poor students. Grade-point averages were computed for each individual by assigning numerical values of 4, 3, 2, 1, and 0 to grades of A, B, C, D, and F, respectively; multiplying each value by its corresponding number of credit-hours; and obtaining the weighted average for all courses. Six distributions were compiled, one for men and one for women within each of the three colleges. The top twenty-seven per cent of each distribution constituted the high-scholarship group and the bottom twenty-seven per cent the low-scholarship group.

A two-by-five contingency table was constructed for each item separately for men and women to determine how well the item discriminated between the good and the poor students. For the men, forty-nine items yielded chi squares significant beyond the .05 level; for the women, twenty items proved significant at the .05 level. Because of this sex difference, separate scoring keys were constructed for men and women.

For the men, two keys were constructed, Form M_1 consisting of forty-two items and Form M_2 containing twenty-nine items. Five of the forty-nine significant items had been eliminated from further consideration because they proved too confusing to eight or more students.[1] The other two were rejected because they exhibited serious inconsistencies of pattern when the responses of the two criterion-groups were compared. The 29-item key for men (Form M_2) included only those significant items for which the response-patterns of the good and poor groups were skewed in different directions; i.e., one group toward the 'almost always' and the other

---

[1] Students were encouraged to mark unclear statements while completing the inventory. Analysis of these data revealed that a small number of items were checked by eight or more students as ambiguous, the remaining items apparently being clear to everyone.

TABLE I.—CORRELATION BETWEEN INVENTORY SCORES AND ONE-SEMESTER GRADES FOR COLLEGE FRESHMEN, MAY, 1951, ADMINISTRATION

College	Men		Women	
	N	r	N	r
University of Texas	138	.46	96	.42
San Antonio College	93	.56	60	.47
Trinity University	54	.54	53	.54
Total	285	.50	209	.46

toward the 'rarely' end of the scale. This minimized the likelihood of chance factors operating in the selection of items for inclusion in the second key. Only one 18-item key (Form F) was constructed for women. Two of the twenty significant items had been eliminated from further consideration in their present form because they contained words found to be incomprehensible to eight or more of the subjects.

As a tentative check on the validity of the 102-item questionnaire, correlation coefficients were computed between grade-point averages and scores on both M_1 and M_2 or on F, for all men and women separately in each college. The correlation between the two keys, M_1 and M_2, proved to be .98. Since they both yielded identical coefficients (.50) when correlated with criterion scores, only the results for M_2, the shorter key, are presented in Table I, together with the results for F, the women's key.

The consistently satisfactory correlations from one college to the next for both men and women—none is below .42—indicate a provisional validity sufficiently high to warrant additional research concerned with its cross-validation on new samples of college freshmen. Such further research was deemed necessary for another reason, too. The above students were tested in May, near the close of their first semester at college. This late in the semester a student is usually aware of how well he is doing in his courses, and this might very well influence his responses to the questionnaire. Evidence bearing upon this question is provided by Gordon (*5*). When twenty-seven student nurses were tested upon entering college and then retested one month later, the correlation between scores on Wrenn's *Study-Habits Inventory* and course grades rose from .26 to .58.

TABLE II.—CORRELATION BETWEEN INVENTORY SCORES AND ONE-SEMESTER GRADES FOR ENTERING COLLEGE FRESHMEN AT THE UNIVERSITY OF TEXAS, SEPTEMBER, 1951, ADMINISTRATION

	N	r
Men	219	.50
Women	176	.52

In September, 1951, the 102-item questionnaire was cross-validated on a fresh sample of two hundred and nineteen men and one hundred and seventy-six women who were entering freshmen at The University of Texas. At the end of the Fall semester, grade-point averages were obtained for every student and correlated with scores on the questionnaire. As indicated in Table II, the results of the cross-validation were highly satisfactory, correlations of .50 and .52 being obtained for the men and women, respectively.

To determine the extent to which performance on this inventory of study habits and attitudes reflects intellectual factors as measured by standard scholastic aptitude tests, scores on the ACE and the *Coöperative Inter-American Tests of Reading*, Advanced Level, Form AE, (CIA) were obtained for most of the students tested in September. The intercorrelations of scores on the inventory of study habits and attitudes, the ACE, the CIA, and one-semester grades are presented in Table III.

Correlations between the study-attitudes inventory and the ACE were .21 and .30, while those between the inventory and the CIA were only slightly higher. Clearly, performance on the inventory of study habits and attitudes is largely unrelated to the usual measures of scholastic aptitude. Consequently, it can be expected to make a unique and important contribution to the prediction of college grades.

TABLE III.—INTERCORRELATIONS OF INVENTORY SCORES, THE ACE, THE CIA, AND ONE-SEMESTER GRADES FOR 144 MEN AND 176 WOMEN

Test	Men			Women		
	ACE	CIA	Grades	ACE	CIA	Grades
Study inventory	.30	.39	.59	.21	.43	.52
ACE		.75	.57		.83	.63
CIA			.58			.64

To determine how much better one-semester grades could be predicted by using a combination of the ACE and the inventory rather than either alone, multiple correlation coefficients were computed among the three variables. The obtained correlations were .720 for men and .744 for women, indicating considerable improvement in the prediction of grades. In view of the probable unreliability of a criterion based upon grades for only one semester, these results indicate that the inventory is a highly promising instrument to aid in understanding students with academic difficulties and to provide a basis for counseling.

A final revision of the questionnaire was made to eliminate items serving no useful purpose and to revise the wording of several items which proved confusing to some freshmen. Thirty-four items were discarded outright and seven new items were added, resulting in a 75-item questionnaire hereafter referred to as the *Survey of Study Habits and Attitudes*.[2] None of the items which had proven to be significant discriminators in the previous research were altered in any way. Subsequent to the research herein reported, extensive normative, validity, and reliability data have been obtained for over three thousand college students throughout the country to whom the *Survey of Study Habits and Attitudes* was administered in 1952 and 1953. Additional item analyses were made resulting in a 36-item key for men and 29-item key for women. Scoring has been simplified and special counseling keys have been devised to identify those statements which represent practices or faults that may be contributing to the student's difficulties. The details of this later research are being reported elsewhere (*6*).

SUMMARY AND CONCLUSIONS

Although a large number of how-to-study manuals and a few study-habits inventories have been published, none thus far have shown any appreciable relationship to subsequent academic performance. The research herein reported was designed to develop a self-rating questionnaire that would measure a student's study habits and attitudes of importance in scholastic success. Analysis of the results obtained for a 188-item preliminary questionnaire administered to high- and low-scholarship groups, matched on relevant variables, revealed that an inventory heavily loaded with statements referring to attitudes and motivation would be superior to the usual study-habits inventory.

---

[2] Test booklets, answer sheets, scoring keys, and manual for the Brown-Holtzman *Survey of Study Habits and Attitudes* may be obtained directly from the publisher, The Psychological Corporation, 522 Fifth Avenue, New York, New York.

A revised questionnaire was administered to two hundred and eighty-five men and two hundred and nine women who were first-semester freshmen in three separate colleges. Using one-semester grades as a criterion, separate scoring keys were developed for men and women. The questionnaire was cross-validated on a fresh sample of two hundred and nineteen men and one hundred and seventy-six women, yielding highly satisfactory results, correlations of .50 and .52 with grades at the end of the semester. A final revision of the questionnaire, referred to as the *Survey of Study Habits and Attitudes*, has since been administered to several thousand students with equally satisfactory results and is now available for general use.

From these studies the following conclusions can be made:

1) Attitudes toward studying can be measured by objective procedures and play a substantial rôle in subsequent academic achievement.

2) Performance on the *Survey of Study Habits and Attitudes* is only slightly related to scholastic aptitude as measured by the ACE *Psychological Examination* or similar tests.

3) The unique predictive validity of the *Survey of Study Habits and Attitudes* is important evidence of its relevance for counseling purposes, diagnostic testing, investigations of the education process, and as a teaching aid in remedial or how-to-study classes.

REFERENCES

1) F. D. Brooks and J. C. Heston. "The validity of items in a study habits inventory," *Journal of Educational Psychology*, 36: 257–270, 1945.

2) C. W. Brown. "The study habits of failing and successful students in the first two years of college," *Journal of Experimental Education*, 9: 205–208, 1941.

3) N. B. Cuff. "Study habits in grades four to twelve," *Journal of Educational Psychology*, 28: 295–301, 1937.

4) A. C. Eurich. "An analysis of self-ratings on studiousness traits," *Journal of Applied Psychology*, 14: 577–591, 1930.

5) H. P. Gordon. "Study habit inventory scores and scholarship," *Journal of Applied Psychology*, 25: 101–107, 1941.

6) W. H. Holtzman, W. F. Brown and W. G. Farquhar. "The Survey of

Study Habits and Attitudes: A new instrument for the prediction of scholastic success," *Educational and Psychological Measurement*, 14: 726–732, 1954.

7) S. R. Laycock and D. H. Russell. "An analysis of thirty-eight how-to-study manuals," *School Review*, 49: 370–379, 1941.

8) N. M. Locke. "The student skills inventory: a study habits test," *Journal of Applied Psychology*, 24: 493–504, 1940.

9) C. O. Mathews and Nora Toepfer. "Comparison of principles and practices of study," *School Review*, 44: 184–192, 1936.

10) W. B. Michael and D. E. Reeder. "The development and validation of a preliminary form of a study-habits inventory," *Educational and Psychological Measurement*, 12: 236–247, 1952.

11) R. C. Myers and D. G. Schultz. "Predicting academic achievement with a new attitude-interest questionnaire—I," *Educational and Psychological Measurement*, 10: 654–663, 1950.

12) C. W. Reeder. "Study habits," *School and Society*, 42: 413–415, 1935.

13) G. M. Schlesser and C. W. Young. "Study and work habits," *School Review*, 53: 85–89, 1945.

14) D. G. Schultz and B. F. Green, Jr. "Predicting academic achievement with a new attitude-interest questionnaire—II," *Educational and Psychological Measurement*, 13: 54–64, 1953.

15) P. M. Symonds. "Methods of investigation of study habits," *School and Society*, 26: 145–152, 1926.

16) H. M. Tiebout. "The misnamed lazy student," *Educational Record*, 24: 113–129, 1943.

17) C. G. Wrenn and R. B. McKeown. *Manual of Directions for Study-Habits Inventory*. Stanford: Stanford University Press, 1933.

18) C. G. Wrenn and W. J. Humber. "Study habits associated with high and low scholarship," *Journal of Educational Psychology*, 32: 611–616, 1941.

ORGANISMIC GROWTH: SEXUAL MATURITY AND PROGRESS IN READING

FRED T. TYLER

University of California

According to the organismic concept, all features of the changing structures and functions of the organism are intertwined in some kind of systematic relatedness. It is claimed, for example, that progress in reading is related to various features of the 'total growth process'. Specifically, it has been suggested that the growth of certain primary sex characteristics is related to progress in reading.

Olson and Robinson have both commented on the possibility of a relationship between sexual maturing and achievement in reading. Olson (*3*, p. 236) observed that boys who are retarded in reading may also be retarded in sexual growth, specifically in the descent of the testes. Robinson (*4*, p. 150) tends to agree with this possibility. Later, Olson (*3*, p. 133-34) expressed the opinion that there is a relationship between achievement in reading and the age of the onset of the spurt in the growth of the testes. At the same time he indicated an interest in further research to determine more precisely the degree and the nature of the relationship.

Stolz and Stolz (*6*, p. 323) failed to find a coincidental relationship between the timing of the onset of accelerated growth of the testes and of height. However, they did observe that for each of sixty-one boys, the spurt in growth of the former occurred prior to the puberal growth spurt of the latter. (Note: The age of the onset of the spurt in the growth of the testes will henceforth be designated as T_1.) The discrepancy between the two ages varied between .75 and 1.25 years for almost eighty per cent of the subjects. There might, therefore, be a sequential rather than a coincidental relationship between these two ages; the correlation of .857 reported by Stolz and Stolz indicates a "strong tendency for these growth phenomena to occur in sequential order" (*6*, p. 344).

Since T_1 was not found to be coincident with a growth spurt in a major physical dimension such as height, we would probably not expect to find it coincident with changes in a functional measure such as reading competence. There remains, however, the possibility of a sequential relationship of the type suggested by Stolz and Stolz.

The specific questions studied for this report are:

1) Is there a marked change in the direction of the reading curve at T_1?

2) Or is there a sequential patterning of the changes in these two phenomena?

READING CURVES OF INDIVIDUAL BOYS

In some longitudinal studies various measurements have been obtained at specified chronological ages. Thus, physical measurements and certain psychological data, such as intelligence scores, may have been gathered within a narrow interval of time centering on each anniversary of the birth date. However, for various reasons, it has often been impossible to give achievement tests at specified ages; such tests are more likely to be administered to all subjects on the same day, so that the results have reference to different chronological ages. Under these circumstances certain adjustments in the data are necessary in order to render them more suitable for group analysis. The nature of one of these modifications (*5*) will be illustrated by means of the data for three boys in the California Adolescent Study.

Chronological ages, reading ages, increments, and increments per year for these boys are shown in Table 1. Their reading ages (RA) are graphed in Figure 1(A). The chronological age at the end of the interval (*6*, p. 340) that included T_1 has been indicated by the vertical arrow. Three different relationships are apparent. In one case the largest annual increase in reading age occurred before T_1,

TABLE 1.—Reading Ages and Annual Ratio of Increments for Three Boys in the California Adolescent Study

Case	T_1	First Test		Second Test		Third Test		Fourth Test		Increments in RA			Annual Rate of Increment in RA		
		CA_1	RA_1	CA_2	RA_2	CA_3	RA_3	CA_4	RA_4	$RA_2 - RA_1$	$RA_3 - RA_2$	$RA_4 - RA_3$	$\frac{RA_2 - RA_1}{CA_2 - CA_1}$	$\frac{RA_3 - RA_2}{CA_3 - CA_2}$	$\frac{RA_4 - RA_3}{CA_4 - CA_3}$
A	145	133	150	144	159	157	170	168	182	9	11	12	10	10	13
B	157	139	153	150	175	162	179	174	192	22	4	13	24	4	13
C	156	143	135	155	137	167	155	179	168	9	18	13	2	18	13

in another after T_1. In the third subject, there were no marked differences among the annual increments.

Certain types of comparisons between these graphs are difficult because the T_1's occurred at widely varying CA's. It was decided to plot the data with a different unit on the horizontal axis, one which would make visual comparisons and statistical analyses more feasible. The new unit refers to the difference between T_1 and the CA's at which the achievement tests were administered. The units were computed from the formula

$$d_i = CA_i - T_1 .$$

(Note: i may be 1, 2, 3, or 4, since the reading tests were administered at four different ages.)

CA_i refers to the chronological age at which the i-th reading test had been administered. Thus, if the first test (i = 1) had been given when a boy was 145 months old, and if T_1 was 140 months, then d_1 in this instance at the time of this testing was 145 − 140 = 5 months. The new units (d) for the three boys whose reading curves are shown in Figure 1(A) are in the d_i column of Table 2; these

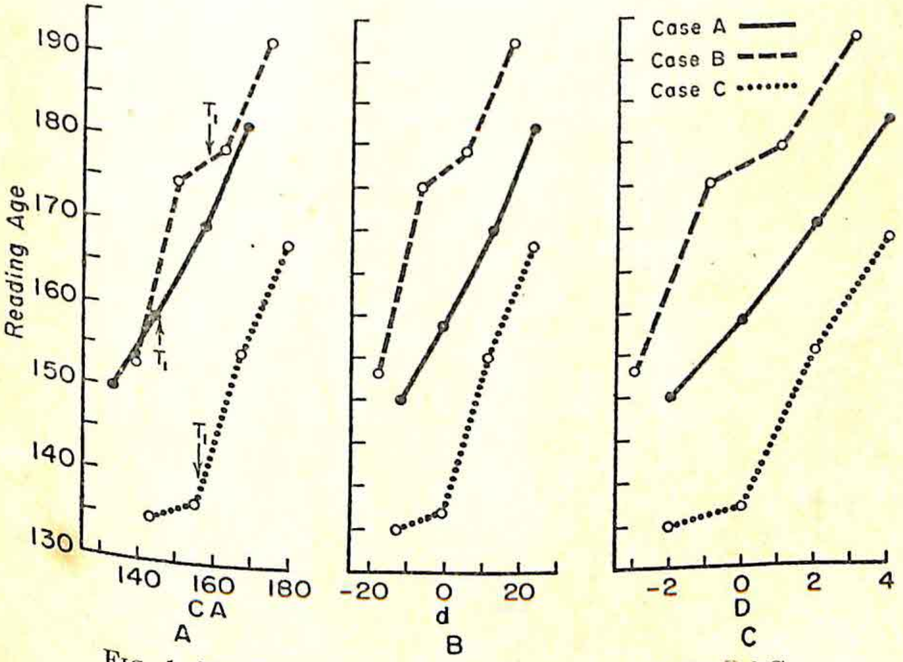


FIG. 1. Reading profiles of three boys, Cases A, B, and C

TABLE 2.—COMPUTATION OF VARIOUS TYPES OF UNITS

Case	T_1	CA_i*				d_i**				D_i†				D_i‡			
		CA_1	CA_2	CA_3	CA_4	d_1	d_2	d_3	d_4	D_1	D_2	D_3	D_4	D_1	D_2	D_3	D_4
A	145	133	144	157	168	−12	−1	12	23	−2	0	2	4	−2	0	2	4
B	157	139	150	162	174	−18	−7	5	17	−3	−1	1	3	−3	−1	1	3
C	156	143	155	167	179	−13	−1	11	23	−2	0	2	4	−2	0	2	4

* CA at each of the four administrations of the reading tests.

** $d_i = CA_i - T_1$.

† $D_i = CA_i - T_1$, with the deviation classified according to Table 3.

‡ $D_i = CA_i - T_1$, with the deviation classified according to Table 4.

units were replotted against RA in Figure 1(B). A comparison of these two figures reveals that this type of presentation makes it easier to generalize the changes in reading ages as a boy passes through T_1 as the point of reference. Two other methods of handling the data will next be illustrated in connection with the statistical analysis of the group data.

STATISTICAL ANALYSIS

The problems of analysis are still not completely resolved since reading ages were not obtained at equal deviations of time from T_1. The d's were consequently changed to D's. The procedure is essentially that of putting the d's into step-intervals according to the scheme shown in Table 3. For example, the D unit is 4 when the difference between T_1 and the age at which a reading test was administered varies from nineteen to twenty-four months inclusive. The curves for the three boys are plotted in Figure 1(C), with the

TABLE 3.—DEFINITION OF THE D UNITS

$d_i = CA_i - T_1$ (months)	D
19 to 24	4
13 to 18	3
7 to 12	2
1 to 6	1
0 to −5	0
−6 to −11	−1
−12 to −17	−2
−18 to −23	−3

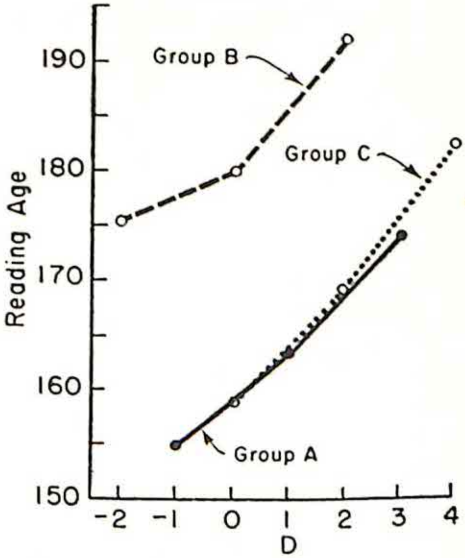


FIG. 2. Mean reading ages of three groups of boys, chronological ages expressed in D units.

D units on the horizontal axis. (The determination of the D units for the three cases discussed above is illustrated in column D of Table 2.)

The values of the means and sigmas of the reading ages of the boys at various D's were then computed. Reading ages were available for ten subjects whose D units were −1, 1, and 3, (Group A), for seven with units of −2, 0, and 2, (Group B), and for twenty with units of 0, 2, and 4 (Group C). (There was no overlap in membership of the three groups.) The means for each group are shown in Figure 2.

There is a slight tendency for the annual increments in reading age to be greater during those intervals that are later in time than during those periods near T_1. For six of the ten subjects in Group A (Figure 2) increments in test scores were larger between +3 and +1 than between −1 and +1. For five of the seven boys in group B, larger increments were found between 0 and +2 than between +2 and +4.

It was decided to choose a second range for the deviations between CA_i and T_1 in the hope of obtaining somewhat larger groups. Consequently, the following units for *D*, having the merit that T_1 is probably more accurately defined as the middle than as the end of an interval, were adopted.

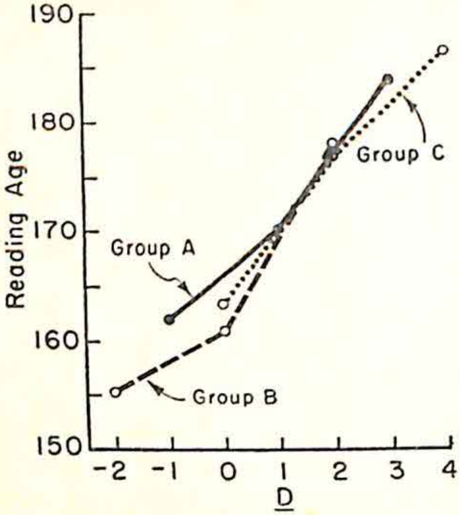


FIG. 3. Mean reading ages of three groups of boys, chronological ages expressed in D units.

Again three groups were located: twenty-three in Group A (D's of -1, 1, and 3), five in Group B (D's of -2, 0, and 2), and fourteen in Group C (D's of 0, 2, and 4). (As in the previous analysis, a boy appears in only one of the three groups.)

The mean RA's at each of the indicated D's are illustrated graphically in Figure 3. For the subjects of groups A and B, the mean annual increment in RA was greater in the period following than during the interval including T_1. In Group C the annual increment in RA was larger in the period immediately following T_1 than in the subsequent period. There does not appear to be any discernible coincidental relationship between differential increase in reading age and T_1. The trend seems to be for T_1 to precede the interval that includes the largest increment in RA.

The reading quotients (RQ) were also obtained and averaged as already described for reading ages. The means are tabulated below:

	N = 23			N = 14			N = 5		
D	-1	1	3	0	2	4	-2	0	2
RQ (Mean)	114	111	112	116	116	113	112	108	108

Clearly the rate of change in reading performance, as measured by the reading quotient, has no organismic relationship to T_1. The

TABLE 4.—DEFINITION OF THE *D* UNITS

$d_i = CA_i - T_1$ (months)	*D*
21 to 26	4
15 to 20	3
9 to 14	2
3 to 8	1
−3 to 2	0
−4 to −9	−1
−10 to −15	−2
−16 to −21	−3

difference between the mean RQ's at *D*'s (−1) and (1) was not significant at the .01 level of confidence. For twenty-eight boys, the largest increases in RA were associated with smaller RQ's than they had earned at some earlier age. For twenty-seven cases, the maximum increase in RA was associated with the largest reported RQ.[1]

CORRELATIONAL ANALYSIS

Stolz and Stolz (*6*, p. 344) reported a correlation of .857 between T_1 and the age of onset of the spurt in growth in height. They also found that in every case T_1 preceded the age at which the spurt in height occurred. A similar type of analysis of the relationship between T_1 and a spurt in reading during adolescence is difficult to duplicate since reading tests were given on only four occasions. However, the increments in reading age were computed and adjusted to indicate the annual rate of increase (as in Table 1). The chronological age at the end of the interval that showed the largest estimated annual change was taken as the age of maximum growth in reading (designated as MGA_R). This procedure is consistent with the method used for determining T_1 (*6*, p. 340). The Pearsonian correlation between MGA_R and T_1 was found to be .267, which is significant at about the .05 level (N = 57).

MGA_R was also computed by defining it as the age at the end of the first interval that showed an annual rate of increase of at least six months of reading age. Actually this procedure changed MGA_R for only four cases. For these data, the correlation between

---

[1] The reading ages were accepted at face value; that is, no account was taken of the fact that small differences might not be reliable.

MGA_R and T_1 was found to be .324, which is significant at the .02 level.

The coefficients reported are consistent with results from other studies of the relationship between physiological changes and behavior (*1*, p. 600). They are not, however, large enough to warrant the conclusion that there is a high degree of relatedness between sexual maturity and such a behavioral characteristic as reading.

SUMMARY AND CONCLUSIONS

Reading ages and quotients and the age of the onset of the spurt of growth in the testes (T_1) were available for fifty-seven boys in the California Adolescent Study. The data were analyzed to provide answers to two questions.

1) Do the curves for reading age and reading quotient improve markedly at T_1, the time of the spurt in growth of the testes?

2) Is there a significant correlation between T_1 and the age of the maximum annual increment in reading age?

On the basis of the data presented, the answer to each question is negative, although it is true that the largest increase in reading ages in the years studied occurred after T_1. There is little reason for believing that there are cyclical changes in reading performance during adolescence. Variations in annual increments were found, but these were in general rather minor and quite irregular. Regarding the second question, neither the theoretical implications nor the practical values of a coefficient of correlation of the order here reported would be of much importance.

No doubt there are important relationships among T_1 and certain other aspects of growth or development or learning. These related characteristics are more likely to be in the realm of physical growth, and possibly in social and emotional learning, rather than in academic learning (except insofar as progress in the latter is related to social and emotional factors). Status respecting testicular growth in adolescent boys seems to be at most only remotely related to progress in reading. The surest guarantee of progress in reading probably lies in the quality of instruction; reading, both developmental and remedial, is dependent upon other characteristics than progress toward sexual maturity.

REFERENCES

1) L. Carmichael (Ed.). *Manual of Child Psychology*. New York: Wiley, 1946.

2) W. S. Gray (Ed.). *Reading and Pupil Development*. Supplementary Educational Monographs. Chicago: University of Chicago Press, 1940.

3) W. C. Olson. *Child Development*. Boston: Heath, 1949.

4) N. M. Robinson. *Why Pupils Fail in Reading*. Chicago: University of Chicago Press, 1946.

5) F. K. Shuttleworth. "The physical and mental growth of girls and boys age six to nineteen in relation to age at maximum growth." *Monogr. Soc. Res. Child Developm.*, 1939, 4, No. 3.

6) H. R. Stolz and L. M. Stolz. *Somatic Development in Adolescent Boys*. New York: Macmillan, 1951.

SELF-APPRAISAL IN TEACHER PERSONALITY

SISTER MARY AMATORA, O.S.F.

Research Professor of Psychology

St. Francis College, Fort Wayne, Indiana

Experiments in the study of personality have increased rapidly in the past decade both in kind and in scope. Yet, every careful investigator knows that while considerable progress has been achieved the goal remains far ahead.

Various measuring instruments have been devised for the purpose of studying personality. These may be grouped broadly in three principal categories; namely, objective tests, biographical and life-record data, and the so-called self-rating scales and inventories.

Cattell (*9*, *10*) maintains that any sufficiently broad and important personality trait will show itself almost equally well in any of these three media. He contends that an identifiable factor found in one media may not be strictly the same factor, yet the high correlations will indicate the functioning of the same basic trait.

Nevertheless, self-rating as well as rating by others has certain limitations. In order to do an adequate job of rating, one must achieve a proper perspective. A good judge of personality must be capable of detachment from his social milieu and at the same time be able to view the subject of his rating relative to current social environment. Self-judgment demands a similar self-detachment.

For all practical purposes, however, a frame of reference or viewpoint, consequent upon one's experiences will be present. An individual's evaluation of the self or others is to some extent dependent upon the group with which he makes the comparisons, be it done consciously or unconsciously.

In order to circumvent this limitation Wolff (*23*) developed a technique which he called 'unconscious self-judgments'. His study showed that the individual rarely judges himself to be average. Huntley (*14*), who duplicated Wolff's experiment, found similar results. The unconscious self-ratings yielded extreme judgments.

Another study of self-opinions, that of Sears (*20*), compared self-ratings with the judgments of several intimate friends. He found defense reactions preventing accurate judgments of others in those who were low in self-insight.

Studies on levels of aspiration by Gruen (*13*) report self-ratings to be related to emotional maladjustment. Those among his subjects who were emotionally unstable tended to give either very high estimates or judgments well below actual performance, whereas the well-adjusted were consistent in judging slightly above actual performance.

Maslow (*16*) shows interrelations between security and self-esteem. His studies indicate a general tendency for persons who feel insecure to evaluate themselves poorly. Yet there is no close correlation between security and self-esteem. Stagner (*22*), from a study of five hundred twenty-four college men, concludes that it is normal to have a moderate tendency to over-rate oneself on rating scales and questionnaires.

In general it seems that inadequately integrated individuals are likely to rate themselves either very high or very low, while, on the other hand, moderate self-esteem tends to be correlated with better adjustment.

At best, self-rating is an introspection and as such is subject to the limitations of introspective psychology. It also suffers from the disadvantage of a certain amount of unreliability, which can never be reduced by increasing the number of observers. There is just one 'self'.

Withal, however, Burt (*8*) considers self-rating as second only to life-record methods for accuracy in personality measurement. He further maintains that when used with coöperative groups, self-rating ties for first place in accuracy. With dishonest and ignorant subjects it falls below every other method.

PROCEDURE

The present investigation seeks to ascertain the relationships between the self-ratings and ratings by peers for a group of teachers on a number of personality variables. A modified form of the Kelly (*15*) scale, prepared by the writer for another study (*3*), was used. This consists of thirty-six characteristics or traits related to personality, each of which is rated on a separate scale in values of from one to ten. The ends of the continuum for the separate scales were staggered so as to lessen any possible halo effect.

The population sampled in the study consisted of teachers of grades four through eight in both rural and city schools, in public and private schools.

Self-ratings and ratings by four fellow teachers were secured for each of one hundred teachers. Ratings by colleagues were averaged for each teacher on each scale. These were then compared with the self-ratings of each teacher on each of the thirty-six scales. In order to facilitate interpretation, all inverted ends of the scales were reversed so that for purposes of comparison the desirable ends of the continuum would be at the upper ends of the scales and the less desirable qualities at the lower ends of the scales.

DEGREE OF RELATIONSHIP

Pearsonian r's were computed for each of the personality variables. As will be noted in Table 1, these range from 54 ± .07 to .13 ± .10 with a median r of .34 ± .09. It will be observed first that all correlations are positive showing that the ratings by self and others are in the same direction. It is further observed that twenty-eight of the thirty-six r's are statistically significant at the one per cent level of confidence and that three more are significant at the five per cent level.

The individual personality traits showing the higher degree of relationships and those exhibiting less agreement are worthy of consideration. Lowest agreement is present on such components of personality as physical attractiveness, degree of tolerance, amount of energy and pep displayed, absence of egotism and self-centeredness, and estimates of intelligence. Means on these five scales, though not statistically significant in their differences favored the 'self' in the tolerance, pep, and absence of egotism. On physical attractiveness and intelligence the mean ratings of colleagues exceeded slightly the mean self-ratings. Self-estimates were the more conservative on these two traits.

Glancing at the personality elements showing highest agreement between self-rating and ratings by others, one observes such traits as persistence in sticking to a task, thoughtfulness of others, common sense, a sense of humor, dependability, punctuality, and

TABLE 1.—AGREEMENT BETWEEN SELF-ESTIMATES AND OPINIONS OF COLLEAGUES

Personality Variable	$r \pm$ SE		t values		Personality Variable	$r \pm$ SE		t values	
Persistence	.54	.07	1.00	S*	New social situations	.34	.09	2.19	S
Boring-entertaining	.48	.08	0.35	P†	Courtesy	.32	.09	1.25	P
Thoughtfulness of others	.47	.08	1.27	P	Selfishness	.32	.09	0.87	S
Boisterous-quietness	.45	.08	0.14	S	Neatness	.31	.09	0.47	P
Common sense	.45	.08	0.06	P	Religiousness	.30	.09	0.36	P
Nervous-calmness	.44	.08	0.06	S	Jealousy	.30	.09	2.0[illegible]	S
Sense of humor	.43	.09	2.20	S	Graceful-awkwardness	.30	.09	1.46	P
Dependability	.43	.09	1.56	S	Good-sportsmanship	.28	.09	0.06	S
Punctuality	.42	.08	2.18	S	Voice	.28	.09	0.36	P
Sociability	.39	.09	0.06	P	Coöperation	.25	.10	2.13	S
Temper	.39	.09	0.36	S	Conventionality	.23	.10	1.00	P
Generosity	.37	.09	0.00	—	Sincerity	.21	.10	0.59	S
Disposition	.37	.09	1.13	S	Popularity	.19	.10	0.00	—
Initiative	.37	.09	0.71	S	Physical attractiveness	.18	.10	0.08	P
Culture	.37	.09	0.38	P	Tolerance	.18	.10	1.14	S
Patience	.35	.09	0.06	P	Energy and pep	.17	.10	0.94	S
Interests	.35	.09	‡4.00	S	Egotism	.16	.10	0.47	S
Honesty	.34	.09	2.00	S	Intelligence	.13	.10	1.86	P

All r's at or above .196 are statistically significant at the 5% level.
All r's at or above .255 are statistically significant at the 1% level.
* Signifies favor self.
† Signifies favor peers.
‡ Statistical significance at 1% level.

others. Whether the higher degree of agreement between self-rating and peer-rating on these traits be due to a greater facility in observation or to some other cause is a moot question.

COMPARISON OF MEANS

A comparison of means for these traits was made by the t technique. The second and fourth columns of Table 1 reveal those for self-ratings to exceed those for peer-rating on such qualities of personality as sense of humor, persistence, dependability, punctuality, calm and self-controlled, optimistic disposition, not easily angered, resourceful and having much initiative, generally quiet rather than boisterous, wide range of interests, honest and just, unselfishness, jealousy, good-sportsmanship, coöperation with others, and sincerity.

Though differences in means for these are not great they do corroborate the studies mentioned earlier and also that of Robertson (*19*) in that self-estimates usually exceed slightly the ratings by others. This does not necessarily mean that the individuals are rating themselves too highly. The same psychological factors that tend to lead one to rate oneself slightly higher than one's peers could lead one to rate oneself peers slightly lower than oneself. Similarly would the same factors cause others to rate one slightly lower than one rates oneself. Which is the better criterion is a question to which no adequate answer is available at present.

The remaining traits showed either no difference or a slight difference in favor of the ratings by others.

RELIABILITY OF PEER-RATING

In order to ascertain some measure of the reliability of the peer-ratings, Pearsonian *r*'s were computed for the five elements of personality showing the highest agreement with self-estimates and the five showing the lowest agreement.

Table 2 shows the results of this analysis when the average ratings of two teachers against two teachers randomly selected were computed and stepped-up by the Spearman-Brown formula. A glance at this table indicates no great differences in the reliability of peer-rating. All achieve statistical significance at the one per cent level of confidence. Of the ten traits herein presented, the only one noticeably lower than the others is tolerance.

TABLE 2.—RELIABILITY OF PEER-RATINGS FOR FIVE HIGHEST AND FIVE LOWEST PEER-SELF CORRELATIONS

Five Variables with Highest *r*'s	Reliability	Five Variables with Lowest *r*'s	Reliability
Persistence	.83	Physical attractiveness	.89
Boring-entertaining	.85	Tolerance	.47
Thoughtfulness of others	.90	Energy and pep	.83
Boisterous-quietness	.91	Egotism	.78
Common sense	.83	Intelligence	.86

CONCLUSION

In conclusion it might be stated that as a group teachers are conservative in their self-estimates on personality variables, that self-ratings and ratings by fellow teachers show a statistically significant degree of relationship on most observable components of personality, and that the findings of the present study do not differ greatly from those of other similar investigations.

True, the final answers to the numerous questions on the measurement of personality are by no means available as yet, but it is hoped that continued research in this vital area will bring nearer the anticipated goal.

REFERENCES

1) M. Amatora. "Boys' personality appraisals differentiate teacher groups." *School and Society*, 76: 184–187, 1952.

2) ——"Children's rating of associates." *Journal of Experimental Education*, 13: 20–22, 1944.

3) ——"Comparability of child and adult personality scales." *Journal of Educational Psychology*, 35: 309–313, 1944.

4) ——"Factor analysis of children's personality scale." *Journal of Psychology*, 18: 197–201, 1944.

5) ——"Reliability of personality scale." *Educational and Psychological Measurement*, 12: 132–136, 1952.

6) ——"The education factor in personality appraisal." *Journal of Experimental Education*, 21: 271–275, 1953.

7) H. E. Brodgen. "A factor analysis of forty character tests." *Psychological Monographs*, 52: 39–56, 1940.

8) C. L. Burt. "The assessment of personality." *British Journal of Educational Psychology*, 15: 107–121, 1945.

9) R. B. Cattell. "The main personality factors in questionnaire, self-estimate material." *Journal of Social Psychology*, 31: 3–38, 1950.

10) ——and D. R. Saunders. "The matching of personality factors in behavior rating, questionnaire, and objective test media." *Journal of Social Psychology*, 31: 243–260, 1950.

11) W. W. Cook and C. H. Leeds. "Measuring the teaching personality." *Educational and Psychological Measurement*, 7: 399–410, 1947.

12) A. E. Ellis. "The validity of personality questionnaries." *Psychological Bulletin*, 43: 385–440, 1946.

13) E. W. Gruen. "Level of aspiration in relation to personality factors in adolescents." *Child Development*, 16: 181–188, 1945.

14) C. W. Huntley. "Judgments of self based upon records of expressive behavior." *Journal of Abnormal & Social Psychology*, 35: 398–427, 1940.

15) E. L. Kelly. "A 36-trait personality rating scale." *Journal of Psychology*, 9: 97–102, 1940.

16) A. H. Maslow, E. Hirsh, M. Stein, and I. Honigmann. "A clinically derived test for measuring psychological security-insecurity." *Journal of General Psychology*, 33: 21–41, 1945.

17) L. L. McQuitty. "Diversity of self-endorsements as a measure of individual differences in personality." *Educational and Psychological Measurement*, 9: 3–14, 1949.

18) M. G. Powell. "Comparisons of self-rating, peer-ratings and experts'-ratings of personality adjustment." *Educational and Psychological Measurement*, 8: 225–234, 1948.

19) A. E. Robertson and E. L. Stromberg. "The agreement between associates' ratings and self-ratings of personality." *School and Society*, 50: 126–127, 1939.

20) R. R. Sears. "Experimental studies of projection. I. Attribution traits." *Journal of Social Psychology*, 7: 151–163, 1936.

21) E. Shen. "The validity of self-estimate." *Journal of Educational Psychology*, 16: 104–107, 1925.

22) R. Stagner. "Wisconsin Scale of Personality Traits." *Journal of Abnormal Social Psychology*, 31: 463–472, 1937.

23) Werner Wolff. *Expression of Personality*. New York: Harper & Brothers. 1943.

ABNORMAL PSYCHOLOGY AS A SELECTIVE FACTOR IN THE COLLEGE CURRICULUM[1]

EUGENE S. MILLS

Whittier College

In recent years psychologists have shown an increasing interest in the problems and opportunities associated with the teaching of their field. Such interest is revealed directly in articles on the teaching of psychology (*3*, *4*, *15*, *17*, *18*) and is implied in many articles dealing with professional problems (*1*, *6*, *11*). Since World War II, during which clinical psychology made such rapid strides, there has also been an increasing concern over the problem of professional personnel in psychology, with particular reference to the process of selecting students for graduate study who give the greatest promise of competence in the field (*8*).

In addition to the above developments, psychologists have continued to devote their attention to problems of student adjustment. Recommendations have been made as to the function of the psychology program in this connection (*2*, *5*, *7*, *10*). Throughout the field there has been a direct interest in what psychology courses can mean to the student and in those teaching techniques which appear most helpful to him.

In spite of the manifest interest in those educational and adjustment problems which are associated with the teaching of psychology, relatively little attention has been given to the possibility that psychology courses may operate selectively in the college curriculum. Most writers seem to have left untreated the rather special nature of the student group which may be enrolled in psychology courses. If it were found that a psychology course (or courses) contained a selected group of students—drawn by the special nature of psychological study—it would appear that the implications for the teaching of psychology would be worth considering.

Because of the nature of the material considered in abnormal psychology, this course presents an interesting opportunity for studying the possibility of the selective function of psychology in the undergraduate curriculum. The morbid nature of the material composing the content of this area of study intrigues a large number of students. While it is undoubtedly safe to assume that the students comprising a class in abnormal psychology constitute a heterogeneous group with respect to personality adjustment, the present study was designed to investigate the following hypothesis:

---

[1] A portion of this paper was read at the meetings of the Western Psychological Association, Seattle, Washington, June, 1953.

The undergraduate course in abnormal psychology operates as a selective factor in the college curriculum, attracting students who reveal, as individuals and as a group, different patterns of personality adjustment than those revealed by students in non-psychology courses in the curriculum.

SUBJECTS

The subjects of the study were chosen from the enrollment in two undergraduate courses in the curriculum of a small liberal arts college. Group A was composed of ten men and eleven women students selected from a course in abnormal psychology according to the following criteria:[2]

1) The student must not be currently enrolled in any other course in psychology, sociology, or criminology.

2) The student must not at that time be undergoing any psychological or medical therapy.

3) The student must have completed at least two years of college.

The subjects in Group A had a mean age of twenty-two years and a median age of twenty-one years.

Group B was composed of eleven men and eleven women students selected from a modern European history course. The criteria for the selection of the subjects in Group B were the same as those for Group A, the only exception being that none of the subjects was enrolled in any psychology course. The mean age for the subjects in Group B was 20.7 years with a median age of twenty years.

---

[2] The criteria for the selection of the subjects were set up in such a manner as to facilitate the study of certain other factors not reported on in the present article.

PROCEDURE

At the beginning of the semester the Rorschach test was administered individually to all subjects in both groups and scored according to the Klopfer system (*9*). In order to derive adjustment ratings which were largely objective and which were sufficiently consistent to lend themselves to intergroup and intragroup comparisons, the Munroe Inspection Technique (*13*) was utilized. A sub-analysis of certain Rorschach scores was made in order to facilitate individual and group comparisons. Of particular importance in this connection was a study of the *erlebnistyp* or introversion-extraversion relationship in each record.

In addition to the above operations, all subjects in both groups were administered a story completion test which was patterned after the Madeleine Thomas Completion Stories Test (*12*) and which consisted of a series of story items dealing with college life. *Typical story items are.—*

a) A student is in college. He writes a letter home. What does he say in the letter?

b) This student answers a question incorrectly in class. In the hall after class another student says to him: "You sure messed that one up, didn't you?" What does he do?

c) The student has been going steady with a certain girl. One night she refuses to go out with him. Why?

The group administration of this test occurred during the second meeting of each of the classes. A theme analysis of the story completion material made possible further individual and group comparisons.

Each subject in Group A and Group B wrote an autobiography at the beginning of the semester. In requesting autobiographical material the experimenter left the situation unstructured, asking only that each subject submit some account of his life. Two methods were used for analysis of the autobiographical material submitted by the subjects of the study. The first was essentially a qualitative appraisal of the autobiographies as background against which other material was viewed.

The second method of analysis yielded quantitative results. In the search for essential differences between groups A and B, the experimenter used the ratings of two outside judges, both of whom were candidates for the doctorate in psychology and had had extensive clinical experience. Each judge was given a rating scale which the writer constructed using several of the personality variables derived from Murray's (*14*) research with college students. A five-point scale was assigned to each of the following variables: (a) intraceptive-extraceptive, (b) endocathection-exocathection, (c) projectivity-objectivity, (d) succorance-nurturance, (e) dominance-submission, (f) radical sentiment-conservative sentiment, (g) emotionality-placidity, (h) self-love—object-love, and (i) affiliation-rejection. A final item provided space for comment regarding the judge's impressions of recurring themes revealed in the autobiography. The judges were instructed to adhere closely to the description of the variables provided by Murray (*14*, p. 142), and to use such clinical impressions as they might glean from the autobiographical material.

By assigning arbitrary scores to points on the rating scales it was possible to derive mean ratings for the various scales and thus facilitate comparisons between the groups. A further operation involved computing a total 'deviation score' for each subject by summing deviations from the mean for each scale. Ranking the subjects according to the size of this score made possible a rough correlation of the judge's ratings for each group.

RESULTS

The Rorschach records obtained from the subjects in Group A received a total of three hundred eleven checks, according to the Munroe Inspection Technique, while those in Group B received a total of two hundred twenty-five checks. (The higher figure is suggestive of poorer adjustment.) Table 1 reveals the ratio of the difference between the means of the two groups to be 3.33. This

TABLE 1.—SUMMARY OF MUNROE INSPECTION DATA FOR GROUPS A AND B

	Group A	Group B
Total checks	311	225
Range	8–34	6–16
Mean	14.76	10.22
Median	16.00	11.00
SD	5.67	2.70
t	3.33	

reflects a significant difference in mean adjustment scores between Groups A and B at the beginning of the semester's study.

The M:sum C ratio, important in assessment of introversive-extraversive relationships, was computed for each Rorschach record. While other material must be considered in making a final decision as to introversive or extraversive orientation in life, i.e., the quality of M and the way in which color is used in the responses, it may be noted that eleven Group A and ten Group B Rorschachs were weighted toward the introversive side. Although the number of cases in which the ratio was weighted toward the introversive side is approximately equal for both groups, Group A contained six cases in which the weightings were so extreme as to suggest a marked tendency in this direction. Study of the other test material obtained from these subjects indicated that this marked introversive orientation constituted a major problem for them in adjusting to college life.[3]

Further evidence of the significance of an introversive orientation in these six cases is derived from two judges' rankings of all Rorschach records on the basis of degree of introversion manifested. The judges were instructed to study the total Rorschach picture in each case, utilizing all quantitative and qualitative clues available. A rho of .83 was obtained when these two rankings were correlated. The six Group A cases referred to above were ranked within the top eight by one judge and within the top ten by the other.

A sub-analysis of certain Rorschach scores suggested greater anxiety and insecurity in the Group A subjects. A significantly larger number of m, k, and K responses appeared in the Group A Rorschachs (t of 3.07, 3.16, and 3.21, respectively). Klopfer and Kelly (*9*, p. 242) suggest that anxieties do not have a pronounced rôle in the personality picture unless at least three K or k are present in one record. Study of the data revealed five Group A Rorschachs with three or more K or k responses, while only one such record was obtained from Group B.

---

[3] It is perhaps worth noting that one of these six subjects who received thirty-four Munroe checks at the beginning of the semester, experienced a major psychotic break during the semester and was institutionalized. Another of the six, who received twenty-eight Munroe checks, continued in school during the semester in spite of serious social and academic difficulties climaxed by a panic reaction to a dating experience.

TABLE 2.—FREQUENCY OF OCCURRENCE OF THREE STORY COMPLETION THEMES: ANXIETY, ESCAPE, DEPRESSION

Themes	Group A	Group B	t
Anxiety	29	13	3.32
Escape	24	10	3.27
Depression	29	15	3.30

Although the total Rorschach picture must be considered in order to gain the best evaluation of depressive tendencies, Klopfer and Kelly (*9*, p. 243) have suggested that achromatic responses are indicative of such tendencies if they outnumber bright color responses at least two to one. Eight Group A subjects produced records in which achromatic responses outnumbered the color responses by at least two to one, while such was the case in only two Group B records.

Theme analysis of the responses to the story completion test tends to support the Rorschach findings. Table 2 presents the frequency of occurrence of anxiety, escape, and depression themes in the story completions of Groups A and B. It is apparent from this table that the subjects in Group A produced a significantly larger number of these three themes than did those in Group B. An individual analysis of the story completion tests revealed that seventy per cent of the Group A anxiety themes occurred in the records of six subjects who also produced sixty-four per cent of the escape and eighty-six per cent of the depression themes. Five of these same subjects were to be found in the group of six judged to be markedly introversive on the basis of Rorschach evidence. The Group B anxiety, escape, and depression themes were rather evenly scattered among the subjects.

Study of the test material revealed a high degree of expressed interest in, and concern over, sexual matters on the part of the Group A subjects. This is reflected in the results of the story completion theme analysis in which fifty-six erotic themes appeared in the Group A tests as opposed to thirty for the Group B subjects. Analysis of the Rorschach records for Groups A and B revealed seventeen sex responses in those records obtained from the Group A subjects and only one sex response in the Group B records. The seventeen Group A sex responses occurred in the records obtained from five subjects, four of whom were included in the group of six that produced most of the anxiety, escape, and depression themes on the story completion test.

The differences between the mean ratings on the personality variable scales were statistically significant on only the succorance-nurturance and self-love—object-love scales, with the Group A subjects revealing a tendency toward succorance and self-love. This difference was largely attributable to the ratings of five of the autobiographies. When ranked on the basis of total deviation scores derived from the ratings, a Spearman rho of .81 was obtained between the judges' impressions of Group A autobiographies and a rho of .58 between their impressions of the Group B autobiographies. While the judges were in relatively close agreement with respect to the degree of deviation of the Group A subjects, their evaluation of the Group B autobiographies showed only moderate agreement. It is worth noting in this connection that each judge volunteered to the experimenter that he felt the Group A autobiographical material was, on the whole, much less difficult to rate than the Group B material due to the fact that the subjects in Group A had written in such a manner as to make personal tendencies much more evident.

The above evidence, plus clues found in personal interviews with the Group A subjects, points up a rather important characteristic of the psychology student. The psychology student, due largely, no doubt, to previous experience in psychology courses, reveals greater frankness in expression of problems related to his more intimate relationships. The study of psychology develops in the student an awareness of the importance of psychological relationships and mechanisms, and while he may not have matured psychologically to the point where he has a great amount of insight into his own personality, he is more disposed to search for reasons back of his behavior than is the non-psychology student.

DISCUSSION

While an analysis of the data produced a significant difference between Groups A and B in mean Munroe scores, as well as in certain areas of the story completion and autobiographical material, a subanalysis of the group data suggests that most of the differences between the two groups can be accounted for in terms of the responses of six of the Group A subjects. In light of the hypothesis governing this study it is legitimate to question whether the course in abnormal psychology functioned selectively in these six cases. That is, can it be assumed that the measurable differences between Groups A and B are due to the special function of the abnormal psychology course in the curriculum? In this connection it is instructive to note that, on certain preliminary questionnaire material obtained from all subjects, each of the six deviant subjects in Group A indicated that he had enrolled in the abnormal psychology course in the expectation that the course might lead to better personal adjustment. Such was the expressed motivation in two other Group A subjects. This suggests that the abnormal psychology course did function selectively in the curriculum studied, offering to certain students what appeared to be an opportunity for bettering personal adjustment.

Four of the six psychology students who deviated from the rest of the group in test responses were psychology majors. Each of these four indicated an intention to take graduate work in psychology and enter the field professionally. While one must obviously be cautious in making inferences from such a small group, this result suggests that psychology as a field may operate selectively in the college curriculum with respect to personality adjustment. In this connection it is desirable to question whether many such students actually continue in the field. Further, one might question whether such manifest anxiety and insecurity at this stage of development necessarily constitute an unfavorable background for future competency in the psychological field.

The psychology students who stand out from the rest of the Group A subjects appear, as a group, to be anxious, introverted, and preoccupied with their own personal problems. It seems likely that the succorance and self-love revealed in the study may be concomitant features of this general personality picture. It is not clear as yet to what extent these personality characteristics may be present in other psychology students who reveal marked adjustment problems, or whether further research might disclose any common personality pattern. On the basis of evidence now available, the writer has little faith in the possibility that a unique personality will be found in the psychology student. The present study, however, suggests that a core consisting of anxiety, introversion, and a preoccupation with personal problems may be predisposing factors for an interest in some psychology courses.

The psychology student's greater frankness in expression of problems related to his more intimate relationships, seems a characteristic of the Group A subjects as a whole and represents their major point of commonality. This suggests that the study of psychology does have an effect upon the student which goes deeper than a strictly intellectual experience. From the standpoint of mental health it is likely that this development can be regarded as a positive one in the lives of most students. Such increased frankness in recognizing and discussing one's problems may represent an important first step in solving them. However, that nearly one-third of the Group A subjects displayed marked adjustment difficulties would seem indicative of a condition which may be more prevalent in psychology courses than has been supposed. This condition obviously stands in need of confirmation, although the selective function of the abnormal psychology course studied seems sufficiently apparent in this situation as to demand some consideration.

Regardless of whether or not the psychology instructor wishes to teach the abnormal psychology course as an adjustment class, it is probable that most such classes contain a number of students whose needs and expectations are such that problems of personal adjustment will be foremost in their minds. This points to the need for instruction which is sensitive to the basic concerns of the students involved. It has yet to be demonstrated that such sensitivity on the part of the instructor is incompatible with the demand for sound student scholarship.

SUMMARY

1) Analysis of Rorschach records revealed a significant difference in mean Munroe adjustment scores at the beginning of the semester with the psychology students revealing more signs of poor adjustment.

2) An analysis of Rorschach and story completion material suggested greater anxiety, depression, and expressed interest in sexual matters on the part of the Group A subjects.

3) Judges' ratings of autobiographical material revealed a tendency toward succorance and self-love on the part of the Group A subjects.

4) An analysis of test data suggested the possible significance

of an introversive orientation in the lives of the more poorly adjusted psychology students.

5) The psychology student reveals greater frankness in expression of problems related to his more intimate relationships.

6) Most of the differences between the two groups can be accounted for, in terms of the responses of six of the Group A subjects who displayed serious adjustment problems.

CONCLUSIONS

1) The abnormal psychology course operated as a selective factor in the curriculum.

2) The results of this study suggest the need for similar inquiries within the curricula of other colleges.

REFERENCES

1) APA Committee on Ethical Standards for Psychology. Section 6. "Ethical standards in teaching." *Amer. Psychologist*, 1951, 6, 649–661.

2) F. K. Berrien, "A new type of elementary course." *Amer. Psychologist*, 1947, 2, 148–150.

3) C. E. Buxton. "The introductory psychology course." *Amer. Psychologist*, 1946, 8, 303–310.

4) ——"Teaching: Have your cake and eat it too?" *Amer. Psychologist*, 1951, 6, 111–118.

5) G. F. Castore, and F. K. Berrien. "A student evaluation of a case method course." *Amer. Psychologist*, 1950, 5, 149–151.

6) R. P. Fisher, and R. P. Hinshaw. "The growth of student interest in psychology." *Amer. Psychologist*, 1946, 4, 116–118.

7) D. L. Grumman, and T. Gordon. "The counseling center at the University of Chicago." *Amer. Psychologist*, 1948, 3, 166–171.

8) E. L. Kelly, and D. W. Fiske. *The Prediction of Performance in Clinical Psychology*. Ann Arbor: Univ. of Michigan Press, 1951.

9) B. Klopfer, and D. M. Kelly. *The Rorschach Technique*. Yonkers N.Y. World Book Company, 1942.

10) A. Magaret, and A. C. Sherriffs. "Personality factors influencing the learning of the first course in psychology." *J. gen. Psychol.*, 1947, 37, 66–77.

11) E. S. Mills. "Defining the field of psychology for the student." *Amer. Psychologist*, 1953, 6, 250.

12) ——"The Madeleine Thomas Completion Stories Test." *J. Consult. Psychol.*, 1953, 2, 139–141.

13) Ruth Munroe. "*Adjustment and academic performance of college students*." Stanford University: Stanford University Press, 1945.

14) H. A. Murray. *Explorations in Personality*. New York: Oxford University Press, 1938.

[15] S. L. Pressey. "The place and functions of psychology in undergraduate programs." *Amer. Psychologist*, 1949, 4, 148.

[16] J. W. M. Rothney. "First course in psychology as a guidance project." *J. appl. Psychol.*, 1935, 19, 404.

[17] D. Wolfle. "The sensible organization of courses in psychology." *Amer. Psychologist*, 1947, 10, 439.

[18] D. Wolfle, et al. *Improving Undergraduate Instruction in Psychology.* New York: The Macmillan Co., 1952.

SHORTHAND STUDY AND SPELLING ABILITIES

BARRY T. JENSEN

Sanford University

and

JUNNE W. JENSEN

Fremont Union High School, Sunnyvale, Calif.

"I was a good speller before I studied shorthand." This is a statement commonly heard from commercial students. The purpose of the present investigation was to determine if the study of shorthand negatively affected ability to spell.

PROCEDURE

Two tests of spelling ability were administered to students enrolled in beginning shorthand classes shortly after the opening of school in September. Both tests were again administered to the same pupils about eight months later. Inasmuch as English is a required course and it could be expected to have some effect on spelling ability, a group of students not studying shorthand was tested at the same times. The tests were administered by the regular classroom teacher. Shorthand students in the English classes at the times of testing were excused during the tests.

Analysis was in terms of change in test scores in the total group of shorthand students as compared to change in the control group of non-shorthand pupils. Because other studies have shown relationships between spelling ability and intelligence (*1*, *2*), it was intended that groups matched on the basis of mental ability should be compared. It was determined that such control was unnecessary in this specific case, however.

The experimental group, as indicated above, consisted of students enrolled in beginning shorthand and in eleventh-grade English at Bladensburg (Md.) High School. This group consisted of sixty-nine persons who took both pre- and post-tests. A group of fifty-eight non-shorthand persons enrolled in the same level English classes was used as the control.

TESTS

The two tests of spelling ability used were constructed by the writers especially for this investigation. Test M was a recognition test consisting of forty-eight multiple-choice items. Each item consisted of a short definition and four possible spellings of a word selected from one of several lists of spelling 'demons'. Separate answer sheets were not used—the examinee indicated on the test paper what he thought was the correct spelling. A recognition test was used because of the belief that an important factor in 'good' spelling is being able to recognize an error.

Test M was in two parts, the words of each part being selected randomly from the longer list. The test was considered to be satisfactory in terms of reliability. As indicated in entries one and two of Table 1, estimated reliability was above .80.

Test R was a recall test of forty-eight items. Each item consisted of a phonetic spelling of a 'demon' and a brief definition of the word. The examinees wrote the word as they thought it should be spelled. Separate answer sheets were not used. The test was in two parts of twenty-four items each—the words were selected in the same manner as for Test M. A recall test of this nature was used because the task it presents was thought to be similar to the task of the shorthand writer who spells phonetically and then

TABLE 1.—CORRELATIONS RELATING TO SPELLING TEST M (RECOGNITION) AND SPELLING TEST R (RECALL)

Variables	r	*N*
1) Test M, Part I vs Part II, Corrected S-B	.82	165
2) Test M, Kuder-Richardson Formula 19	.82	80
3) Test R, Part I vs Part II, Corrected S-B	.90	165
4) Test R, Kuder-Richardson Formula 19	.93	80
5) Test M vs Test R (total scores)	.66	165
6) Test M vs Test R (total scores corrected)	.77	165
7) Test M, Pre-test vs End-test	.97	69*
8) Test R, Pre-test vs End-test	.44	69*
9) Ability (CTMM) vs Test M (total score)	.22	65†
10) Ability (CTMM) vs Test R (total score)	.43	65†
11) Ability (CTMM) vs change on Test M	.03	65†
12) Ability (CTMM) vs change on Test R	.01	65‡

* Shorthand students only.
† All available ability scores.

TABLE 2.—MEANS, SIGMAS, AND DIFFERENCES RELATING TO TWO SPELLING TESTS ADMINISTERED TO SHORTHAND STUDENTS (*S*) AND NON-SHORTHAND STUDENTS (*N*)

Test	Group	*N*	Mean	SD	Diff.	P
M, pre-test	S	69	41.76	3.6	1.25	.01
M, end-test	S	69	43.01	2.5		
R, pre-test	S	69	32.86	7.8	2.72	.01
R, end-test	S	69	35.58	7.2		
M, mean gain	S	69	1.25	3.3	.47	Not sig.
M, mean gain	N	58	1.72	4.6		
R, mean gain	S	69	2.72	4.4	.43	Not sig.
R, mean gain	N	58	3.15	5.5		

transcribes into English. As indicated by entries three and four of Table 1, the test had adequate reliability for the purpose.[1]

Both tests were timed; both parts being timed separately. However, the time limits proved to be ample to allow everyone to attempt every item.

The ability measure used in the early stages of analysis (before it was found to be unnecessary) was the California Test of Mental Maturity. It was the latest test recorded on the student's record card in most cases. Some records did not carry an ability score and others had different tests noted. All raw scores had been converted to IQ's when entered on the record cards.

RESULTS

Entries five through ten of Table 1 present data related to the question of how many spelling abilities were being tested. The coefficient of correlation between the tests was less than perfect, the tests were differently related to the same ability measure, and the relationship between first test and re-test differed in the two measures. The correlation between tests was significantly lower than the reliability of either—beyond the .01 level.

It can also be noted that spelling abilities as measured are positively related to intellectual ability, but not highly. However, as is shown by entires eleven and twelve of Table 1, change in spelling test score, the variable with which we are concerned, was practically unrelated to ability.

---

[1] The persons in both groups taking the first test were utilized in obtaining the reliability data—N equalled 165.

Table 2 indicates that the change in spelling score shown by the entire group of shorthand students was statistically significant with regard to both kinds of tests. Because of the zero relationship between ability and change, ability was not taken account of in testing the significance of the differences between the shorthand group and the non-shorthand group. As can be noted from Table 2, the differences between the groups in regard to change in spelling ability were small and not statistically significant.

DISCUSSION

Inasmuch as the relationship between Test M (recognition) and Test R (recall) was too low to permit individual prediction from one to the other, it appears that the possibility of at least two kinds of spelling ability must be recognized. At least in this case, the use of both tests was justified.

It appears from the data presented here, that while the shorthand students made a statistically significant gain in ability during the first eight months they were studying shorthand, this gain cannot be considered as a transfer effect due to the study of shorthand. The same gain was shown by non-shorthand students. It may be that this change was due to the common experience of enrollment in eleventh-grade English, but the present investigation yielded nc data to bear upon that question. It should be considered significant that, contrary to common belief, there was no loss in spelling ability associated with the first eight months of study of shorthand.

Although the correlation between mental ability and spelling ability is not, according to this study, as high as has been reported, this particular result is not at too great variance with the earlier findings. Differences might be explained by the size of the groups concerned and the particular measures used. Of importance, however, is the finding that the small gains in spelling ability are unrelated to mental ability; this finding will be checked by an additional study.

SUMMARY

Tests of two spelling abilities (recall and recognition) were administered to high-school students at the beginning and near the end of the first year of study of shorthand. A control group of non-shorthand students taking the same required course in English took the same tests at the same times. It was noted that at least two somewhat different spelling abilities can be identified. It was further determined that the study of shorthand does not necessarily have a negative effect upon either of the spelling abilities.

REFERENCES

1) Agatha Townsend. "An investigation of certain relationships of spelling with reading and academic aptitude." *Journal of Educational Research*, 1947, 40, 465–471.

2) Edmund G. Williamson. "Mental abilities related to learning to spell." *Psychological Bulletin*, 1933, 30, 743–751.

CONFIGURATIONAL LEARNING IN CHILDREN[1]

ALLEN D. CALVIN

Michigan State College

Although a great deal of work has been done by psychological learning theorists on varying methods of stimuli presentation, the vast majority of this work has been with animals. The present study attempts to point out one of the important stimulus variables operating in a standard learning situation when children are used as subjects. Specifically, the present experiment was designed to test the hypothesis that when stimuli are presented simultaneously to children, those situations in which the stimuli are presented contiguously so that they appear as a single perceptual unit or configuration will be more difficult to master than situations in which the stimuli are discrete units.

APPARATUS AND PROCEDURE

The population for this experiment consisted of thirty-four children enrolled in the nursery school and kindergarten of the University of Minnesota Institute of Child Welfare. Fourteen of the subjects were children from the four-year old nursery group, and the remaining twenty came from the five-year old kindergarten class. The mean IQ for the group was 120. (Due to a research program now underway at the Institute some of the children had been tested with the WISC and some with the Stanford-Binet. Since the IQ of the group was obtained only to give the examiner some rough index of the intellectual level of the subjects, the WISC and Binet scores were combined in order to arrive at a group average.)

The apparatus used in the experiment consisted of the following: A large brown screen, two strips of plastic, one blue and the other black, 5⁄8″ by 4–1⁄4″, two identical yellow cups, and a number of small plastic toys.

The subjects were divided into two groups designated Configurational and Non-configurational, respectively. Each group had

---

[1] The author wishes to acknowledge gratefully the coöperation of Dr. John E. Anderson, Director of the Institute of Child Welfare, University of Minnesota, and Dr. Elizabeth M. Fuller, Principal, Nursery School and Kindergarten of the Institute of Child Welfare, University of Minnesota.

seven children assigned randomly from the nursery school and ten assigned randomly from the kindergarten, making the N for each group seventeen.

Each subject was called in individually and given the following instructions: "We are going to play a game. I am going to hide a prize, like this truck, under one of these cups. If you point to the cup that the prize is under, you get to keep the prize. Would you like to ask any questions?" A large brown screen was then placed between the subject and the two identical cups.

If the subject was in the Non-configurational group, two strips of plastic, one blue and one black, were placed 6″ apart directly in front of the two cups. These colored strips served as the positive and negative discriminanda for the problem. If the subject was in the Configurational group, the two strips of plastic were placed side by side a little in front of and equidistant between the two cups so that they formed a single perceptual unit whose components provided the differential cues for the solution of the problem. A balanced experimental design was used. The stimuli were shifted spatially according to a prearranged random order in which the same spatial arrangement of the stimuli never appeared more than twice in succession. The criterion for solution was set at eight successive correct choices. If a subject had not solved the problem after forty presentations, the trials were terminated. An introspective report was obtained from each subject at the end of the experimental session.

RESULTS AND DISCUSSION

The results obtained were as follows: the mean number of trials for the Non-configurational group was 24.9 and ten of the seventeen subjects solved the problem. The mean number of trials for the Configurational group was 38.7, and only one of the seventeen reached criterion.[2] A Chi-square was computed for solvers vs. non-solvers. Using Yates' correction for continuity a value of 8.6 was obtained which is significant beyond the .01 level of confidence. Thus the hypothesis concerning the relative difficulty of configurational vs. non-configurational presentation of simultaneous stimuli was confirmed.

---

[2] Of the fourteen subjects who came from the nursery school, only two in the Non-configurational group solved the problem and none from the Configurational group mastered it. Of the twenty subjects who came from the kindergarten, eight from the Non-configurational group and one from the Configurational group reached criterion. The problem appears to have been a little too difficult for most of the nursery-school children. Thus if all the subjects had been from the kindergarten group it seems reasonable to assume that the differences obtained would have been even more striking.

Since children, rather than animals, were used as subjects, direct reports could be obtained on the perceptual aspect of the problem. To reproduce these reports verbatim would prove too space-consuming, but in essence they pointed up the difficulty of taking a single perceptual unit and using its components as differential cues. Thus the subjects' introspective analyses also tend to corroborate the objective findings mentioned above.

These findings indicate that in children as well as in animals, the perceptual framework in which the stimuli are presented is often of crucial importance. Koffka (*2*), Kohler (*3*), Tolman (*6*), and other Gestalt and neo-Gestalt psychologists have emphasized this fact which is often overlooked by traditional learning theorists. Bitterman and his associates (*1*), (*4*), (*5*), (*7*), (*8*) have also emphasized the importance of perceptual organization. The author believes that a systematic investigation of this configurational variable as it operates in children's learning would clear up many problems in this area which thus far have failed to yield to conventional experimental attack.

SUMMARY

Thirty-four subjects were divided into two groups. A Configurational group was presented with a simultaneous discrimination problem in which the stimuli to be discriminated were presented as a single perceptual unit. A Non-configurational group was presented with a simultaneous discrimination problem in which the stimuli to be discriminated were presented as discrete perceptual units. The principal hypothesis tested was as follows: when stimuli are presented simultaneously to children, those situations in which the stimuli are presented contiguously so that they appear as a single perceptual unit or configuration will be more difficult to master than situations in which the stimuli are presented as discrete units.

The performance of the Non-configurational group was superior to that of the Configurational group. This fact plus the introspective reports of the subjects was interpreted as giving empirical support to the hypothesis stated above.

REFERENCES

1) M. E. Bitterman, A. D. Calvin, and C. B. Elam. "Perceptual differentiation in the course of nondifferential reinforcement." *J. comp. physiol. Psychol.*, 1953, 46, 393–397.

2) Kurt Koffka. *Principles of Gestalt Psychology*, New York: Harcourt Brace & Company, 1935.

3) Wolfgang Kohler. *The Mentality of Apes*, New York: The Humanities Press Inc., 1951.

4) E. L. Saldanha, and M. E. Bitterman. "Relational learning in the rat." *Amer. J. Psychol.*, 1951, 64, 37–53.

5) D. C. Teas, and M. E. Bitterman. "Perceptual organization in the rat." *Psychol. Rev.*, 1952, 59, 130–140.

6) Edward Chace Tolman. *Purposive Behavior in Animals and Men.* Berkeley: University of California Press, 1949.

7) J. R. Turbeville, A. D. Calvin, and M. E. Bitterman. "Relational and configurational learning in the rat." *Amer. J. Psychol.*, 1952, 65, 424–433.

8) P. Weise, and M. E. Bitterman. "Response-selection in discriminative learning." *Psychol. Rev.*, 1951, 58, 185–195.

BOOK REVIEWS

BRUNO KLOPFER, MARY D. AINSWORTH, WALTER G. KLOPFER AND ROBERT R. HOLT. *Developments in Rorschach Technique. Volume I: Technique and Theory.* Yonkers: World Book Co., 1954, pp. 726.

No book has had more influence on American Rorschach technique—and therefore on clinical diagnostic practice—than the Klopfer-Kelley book of 1942. The present volume is a successor to it in part. A Volume II, on application, is promised. The authors have evidently intended this new book to speak simultaneously to three audiences: to beginners, to advanced practitioners, and to their professional colleagues interested chiefly in research and personality theory.

We have first a handbook for administration and scoring. The didactic chapters are clear, with numerous well-chosen examples. There is a good brief chapter on report writing. Undoubtedly, these chapters fulfill their teaching function. A critic might differ from the authors in some scoring decisions, but in the absence of validation any system for scoring is arbitrary. This manual makes explicit how the Klopfer system is intended to be applied, but of course cannot provide evidence that this set of rules is superior to any alternative.

Throughout the book, a second level is continually evident. These authors are speaking over the heads of students to their professional colleagues, arguing basic issues regarding the Rorschach test. This is especially notable in chapters on interpretation of scores, ratios, sequence, content, and so on. A section on 'theory' covers validation, the relation of Rorschach to contemporary personality theory, and a chapter on Rorschach and ego psychology.

Holt's chapter on personality theories covers briefly the writings of Allport, Rapaport, Angyal, and others, discussing how their theories relate to Rorschach interpretation. Holt succeeds in demonstrating that personality theory makes little contact with Rorschach theory, and the reasons he offers for this are plausible. It was perhaps unwise for him to omit general behavior or perceptual theory, including the works of Miller, Postman, and many others; the relatively mundane and observable constructs of experimental psychology are currently providing an astringent flavor in personality research which Holt's chapter lacks. But his chapter is unthe times of Burrow and what he was after, these sketches can be of value from the point of view of historical growth of their movement as well as illustrating both discoveries and difficulties in searching by way of phyloanalysis for human insights that can have implications for society at large as well as for individual personality growth.

H. MELTZER

Psychological Service Center, St. Louis

W. S. TAYLOR. *Dynamic and Abnormal Psychology.* New York: American Book Company, 1954, pp. 658. $5.50

This book, among other things, attempts to make a special contribution; namely, to present a 'pure' abnormal psychology. It "brings together many of the most important contributions to dynamic and abnormal psychology and organizes them within a framework of concepts of general psychology, developing the concepts according to the data."

For a definition of psychology the author accepts that of William James which he interprets broadly (including behavior and the adjustment of the personality); "dynamic psychology is the psychology of motivation;" and "Abnormal psychology is the psychology of those mental states and of those minds which are relatively far from ideal integration." Apparently we are asked to understand the abnormal in terms of "failures of integration." And we are told that ' "relatively far from ideal integration' is a rough measure, but the most useful measure now available."

By definition the subject matter of abnormal psychology is limited. The subjects dealt with in order in nineteen chapters are as follows: dynamic and abnormal psychology; related fields; historical approach to abnormal psychology; basic concepts; a survey of motivation; special topics of motivation; reactions; connector processes; conflict; dissociation; learning and memory; acquired patterns; the sub-conscious; attention and perception; thought; action and control; suggestion; reactions to stress; and, mental disorders. To these are added an introduction to terms of abnormal psychology, and an index and glossary.

There is a great wealth of material gathered from many sources, with many footnote references, and presented in a very interesting manner. The student is introduced systematically to a wide range of subjects both in general and in abnormal psychology. The reftrate the community implications in the symptomatology of the neurotic or psychotic patient. Illustrative captions for these sketches are: The Calendar of Conformity; Commercial Panaceas; The Constant Error; Everybody's Lonesome; The World and His Wife; Yesterday and Today. The subject matter is presented in nine chapters, illustrative titles of which are: Effect or Emotional Bias; The Social Neurosis; The Social Image; The 'I'-Persona; Man is a Unitary Organism; Tensional Patterns and Man's Behavior.

The point of view illustrated in these sketches is expressed by Thompson in his introduction, as follows: ". . .behavior imbalances, whether expressed in the reactions of the 'normal' or the 'neurotic', are traceable to conflict within the organism's internal tensional patterns and that this physiological disturbance is common throughout our so-called civilized life. There is no longer ground for any group of individuals to maintain that they must be alcoholics-anonymous or neurotics. All of us—individuals and nations—have a *common neurosis*. Man is neurotic." In his foreword, Burrow warns the reader that he will not find in these sketches and essays "unrivaled examples of flawless literary style" and admits that these "vignettes really provide samples of much inept, adolescent thinking," and that they exemplify the various symptoms which the authors were supposedly observing. So what they really do is give the reader who is interested a chance to obtain a glimpse into what Burrow considers to be the first socio-biological clinic. In this volume the reader will find mentioned everything that is identified with Burrow's way of looking at life and human understanding, and the range includes everything from subjective continuity derived from pre-conscious development, which has been called by Burrow and his associates the primary identification, to the more frequently emphasized concept of theirs; namely, the social basis of consciousness. With many of the conclusions and attitudes that Burrow and his disciples express, many people today will agree. The intention of appealing for human understanding is there in these writings and even more so in Burrow's recent publications, but neither in the techniques and certainly not in the sketches do we obtain a kind of warm understanding that could be helpful to very many. To a good many readers without a background of familiarity with Burrow and related writings, many of these pages are likely to be completely unintelligible and strike them as being scholastic rather than enlightening. But to those who are familiar with the literature and erences include Abercrombie (1833), Tuke (1873), Carpenter (1875), Ribot (1882), and Sidis (1898). The richness of material extends to much of the recent studies though not to quite all that might be desired. The valuable collection is also indicated by the well made index and glossary (subjects and names in one alphabetical order) which includes pages 631 through 658.

The question immediately arises as to how a single criterion of abnormality can be accepted as "the best one available," or even as satisfactory. Many will object to the idea that integration and failures of integration are well enough understood or practically useful for the purpose indicated. And many will certainly consider that other criteria generally discussed and recommended for this very difficult problem are still needed and more fundamentally useful; namely, great deviation from statistical norms, very unusual behavior, inadequate and inappropriate responses, nature, number and degree of maladjustments, harmful or disadvantageous actions, behavior that is inappropriate and not in keeping with one's cultural background, behavior that is detrimental or dangerous to self, and behavior that is potentially or actually dangerous to the welfare of others?

Secondly, the question arises as to whether either teachers or students will be satisfied with a text which omits (by virtue of defintion, or perhaps for any other reason) or treats too briefly, subjects that are commonly found in our textbooks on abnormal psychology, such as, the sub- and supernormal, our increasing knowledge of the epilepsies (including EEG results), behavior problems (including EEG results), drug and alcoholic addiction and related problems, the organic psychoses, psychopathy, and psychotherapy. The mental disorders are limited to one chapter (pp. 556–623). Other texts give four to five times as many pages to these disorders.

The critic may wish that more were included in this volume. But he may also consider the possible contribution to science when an author attempts to develop and to present that which belongs to 'pure' science,—here, a pure abnormal psychology. The making of good definitions and the clarification of concepts are assuredly much of the valuable contribution that can be made to science. Perhaps here is an example. But the sciences have not been considerate of those who have made very careful, delimiting definitions of their subjects; sciences have developed far beyond the limits of puristic definitions; they have refused to be fenced in. Abnormal psychology may be found not to be abnormal in this respect.

A book is to be evaluated for what it contributes more than for what it omits. This book contributes much. It has a wealth of material collected through many years, and organized painstakingly with much critical advice and assistance. Some of the definitions and classifications will probably not be accepted. But as a text it will be found to be both very interesting and useful. Omitted topics can be brought in by lecture and by supplementary reading. As a reference book, it will open up the best in the various fields of which it treats.

A. S. Edwards

The University of Georgia

J. Anthony Humphreys and Arthur E. Traxler. *Guidance Services*. Chicago: Science Research Associates, Inc.

Designed for use in the introductory class in guidance, this book suggests guidance services for the high school and college, with a few references to the elementary-school program. It is one of a series of four texts intended to eliminate unnecessary duplication of content in the training of school counselors.

The guidance point of view is defined as an attitude that aims to help the individual toward self-understanding, optimum use of abilities, adjustment to environment, self-decision, and making his unique contribution to society. Guidance is said to fit in with our fundamental beliefs and to answer some of the pressing needs of modern society.

The authors believe that guidance workers should understand the importance of individual differences, the wholeness of personality, and the self-concept and emotions as they affect learning. A historical sketch reviews some of the early guidance programs in public schools and colleges, the impact of the vocational guidance and testing movements, the work of individuals and national organizations, and developments in six other countries. Although some interesting facts are added to previously published histories, noteworthy contributions of psychology, social case work, psychiatry and the Mental Hygiene and Child Guidance Associations are overlooked. Claims made for the adequacy of elementary-school guidance programs are unsupported and somewhat exaggerated.

Tests are named and some are discussed; hints on records and testing practices are given. Group guidance is described in relation to orientation to school, occupational information, discussion of student problems, study skills information and extra-class activities. Weaknesses and possibilities in the homeroom program are mentioned.

Placement, research, evaluation, follow-up, selection and training of staff, trends and predictions, organization of the guidance program are among the subjects treated. Although group therapy, group leadership, remedial services and personal adjustment counseling are among services recommended by the authors, 'how-to-do-it' information in these areas is scanty in this and other volumes in the series. Orientation and inservice training are suggested, but not thoroughly explored. Educational and vocational guidance and test interpretation are more specifically explained.

Current trends in personnel work are reflected in the authors' student-centered point of view, mention of newer services such as group therapy, and recognition of psychological and sociological factors in student problems. They present an overview of the guidance program and provide a list of some supplementary readings on each subject.

MAURICE D. AND JEANNE A. WOOLF

Kansas State College, Manhattan

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume 46 March, 1955 Number 3

A SECOND STUDY OF CHARACTERISTICS OF GOOD AND POOR SPELLERS

DAVID H. RUSSELL

University of California, Berkeley

Stephen Leacock, the Canadian humorist, once said that "People look on spelling as one of the troubles of childhood, like measles and Sunday School and having to obey Father." Since spelling is still a 'trouble' to children, the causes of this phenomenon are of interest.

The first study of characteristics of good and poor spellers by the present writer summarized pertinent research in spelling under the last two of four heads: (a) What words shall be taught? (b) At what grade levels shall these be presented? (c) What are the best methods of teaching and learning to spell? and (d) What characteristics of the individual child affect spelling ability? In the fifteen years that have elapsed since the first report (*36*), some advances have been made in the investigation of each of these four areas but, with the possible exception of the first, the progress has not been notable. The present account reviews briefly some of the problems studied during this time and presents further evidence in the fourth category concerned with the individual learner.

WORD LISTS

The pioneer work of Thorndike, Horn and others in obtaining measures of frequency of use as aids to creating a spelling list has been continued in the last fifteen years. The Thorndike-Lorge list (*46*) has been extended to include thirty thousand words. Rinsland (*35*) obtained the most complete record of words used in children's writing so far available. Dolch (*11*) derived the two thousand 'commonest words for spelling' from various vocabulary studies. Fitzgerald's (*13*) list of 2,650 words included a shorter list of four hundred and forty-nine which he believes contains about three fourths of the words children ordinarily write. Lorge's (*32*) semantic count giving the frequency of meaning of five hundred and seventy multi-meaning words is a valuable resource that has not yet been fully utilized in the preparation of spelling texts and curriculum guides.

Somewhat more specialized lists have been presented by different research workers. Not all of these are concerned with children's written language but they have implications for spelling lists or texts. Kyte (*30, 31*) compiled two lists of one hundred and of five hundred and one useful words derived from other lists and judged highly useful in language arts activities, especially spelling. Stone (*44*) revised his vocabulary list in reading for primary grades. Dale (*10*) compiled a list of seventy-five studies of various types of vocabulary in use and added thirty-seven references on vocabulary grade placement. Stauffer (*42*) listed the prefixes found in Thorndike's *Teacher's Word Book of 20,000 Words* and found that fifteen prefixes accounted for eighty-two per cent of the total number of prefixes. Thorndike (*45*) also presented the commonest suffixes in English words of value in spelling instruction as well as in reading instruction.

The careful word counts of the above studies were adapted to school use with the discovery, first by Ayres, and in more detail by Horn (*26*) and others that three thousand to four thousand of these commonest words made up from ninety-five to ninety-eight per cent of all words children ordinarily write. For some time, this finding seemed to create a clear directive for curriculum-makers and authors of spelling texts: Simply teach from three to four thousand of the commonest words and you will have taught children most of the words they need to know. Unfortunately, perhaps, this solution did not work out so easily in practice for it did not consider the many special words children want to write which do not appear on master-lists; it did not say when the basic words were to be taught; and it did not take account of the abilities and disabilities which help or hinder individual children in learning to spell. We now know beyond reasonable doubt the commonest words in English, but all the problems of the spelling program are not thereby settled.

GRADE PLACEMENT OF WORDS

The placement of words at different grade levels has usually been determined in a school's course of study or a series of spelling texts on a multiple basis. Some of the usual criteria are: (1) Words easy to spell are placed in the lower grades, (2) The most frequently used words appear early in the grades, (3) Spelling words should be related to words used in related language arts, social studies and other curricular activities, (4) Common words used by adults more than children should be placed in the upper grades, (5) Words should be taught, if possible, in the first grade in which children write them.

Perhaps the multiple criteria used by different authors have been responsible for the findings in two studies by Betts (*5*, *6*) that authors of spelling texts do not agree on the grade placement of words. In the first investigation involving 8,645 words in eighteen series of texts, Betts found that only one word was placed at the same grade level by all authors. In the second study, there was agreement in the grade placement of sixty-five of the four hundred and eighty-three words common to eight series of spellers. Fifty-five of these words were placed in the second grade. The disagreement found by Betts is not as great as it seems since many words were placed at adjacent grade levels. Indeed, the findings may be interpreted as desirable in the sense that different school systems or schools should place words earlier or later than others, depending on the children's language backgrounds and other factors such as use.

Even the small amount of agreement found by Betts is complicated further by the relationships between word frequency and word difficulty and by the fact that the most adequate grade placement leaves out of the spelling list many words that children of a certain grade may need to write. Research results are not clear-cut in regard to the first relationship. For example, as early as 1918 Hollingworth (*25*) showed that word meaning was a factor in spelling difficulty, but Wesman and Seashore (*48*) have found that frequency and ease of meaning are not identical. Several writers have shown that the factors other than frequency of use and meaning which made a word difficult to spell include such characteristics as length of word, vowel difficulties and unusual combinations of letters. Gates (*17*) has indicated the difficult spots in a list of 3,876 common words and Spache (*40*) has summarized the research on types of spelling errors. Studies such as these help define the first criterion given above as to which words are easy to spell and which more difficult and may, therefore, be of use in constructing word lists for different grade levels.

In regard to the second complication, in a little study which needs further verification at other levels, Wilson (*50*) found that only fourteen per cent of the different words written by fifth-graders in letters, compositions, original stories and other school writing were to be found in their state spelling text. Curtis and Dolch (*9*) reported that most of the learning of spelling in one Illinois school system was done either before the year in which the words were taught from a text or after the year in which they were taught. Hildreth (*24*) has criticized the lack of usefulness of most spelling texts and prepared lists. She said: "Spellers in common use today in Grade II through Grade VIII are heavily loaded with spelling-contest words typical of our grandparents' day—words that must be studied by children some years ahead of the infrequent times they will ever use them." (p. 261) Horn (*28*), however, has sensibly stated that the issue is not an 'either-or' one of spelling list versus individual study but of some combination of the two approaches. It seems reasonable to deduce that the usual spelling list in a course of study or spelling text does not cover nearly all the child's spelling needs and should be supplemented by locally prepared lists and planned instruction in spelling related to other curricular activities and to writing needs.

METHODS OF TEACHING AND LEARNING TO SPELL

The writing and research on methods of spelling instruction has broken little new ground during the last fifteen years. Two books, Dolch's *Better Spelling* (*11*) and Fitzgerald's *Teaching of Spelling* (*14*) contained many specific suggestions for instruction, some of which have well-established research backing. Wilson (*49*) traced the historical development of spelling instruction through three main periods from 1647 to the present, and noted current trends toward individualized teaching and toward achieving a closer relationship between spelling and other curricular activities. Artley (*1*) reviewed some thirty references in deriving five principles of classroom instruction in spelling. These included provisions for individual differences, need for expression as motivation for learning to spell, the necessity of direct instruction in spelling, the need of the child to become independent in spelling ability, and the desirability of favorable attitudes toward correct spelling.

One fresher approach to spelling instruction has developed in an emphasis upon the relationship of spelling to the other language arts. Russell (*38*), Sparrow (*41*) and Townsend (*47*) have all found rather close relationships between spelling ability and a group of other language abilities. Artley (*2*) and Hildreth (*23*) have reviewed the research on interrelationships of the various language arts. Betts (*5*), Gates (*18*) and Yoakam (*52*) have warned that although spelling and reading are closely related in many ways they are also very different in some characteristics and purposes and may actually interfere with one another.

Another new approach to the problem of better spelling instruction has developed within the area of analysis of spelling ability and disability. One attempt to study the composition of spelling ability by factor analysis was hampered by too narrow a definition of spelling and related abilities (*29*). A more comprehensive view of spelling ability was developed by Nichols (*34*) in a series of tests which included spelling achievement, proof-reading, word meaning, handwriting, visual discrimination and auditory discrimination. A number of studies have also continued the attempt to improve spelling instruction through the analysis of errors made by children. After reviewing some thirty studies, Spache (*40*) systematized a method of recording spelling errors and gave the expected frequency of the different types of errors. Wolff (*51*) compared errors made in written assignments, in regular weekly spelling tests, and on a standardized test in one fifth-grade class. Mechanical errors appear most frequently in essay-type materials due to difficulties in handwriting and rules of English usage. Phonetic errors centered on certain hard spots in words; non-phonetic errors were associated with lack of knowledge of the word.

Further analytical data on how children learn to spell were contributed by the Gilberts (*20*) in studies of eye movements. As in an earlier study (*19*), they found wide individual differences in the way children perceived words when learning to spell them. Combinations of letters which were difficult for some children were easily mastered by others. Poor spellers took longer and looked at words irregularly; good spellers spent more time on the hard spots in words. Limiting the time of study and helping a pupil determine his own needs were effective techniques. The Gilberts' results suggested that increasing amounts of the same study procedures will not improve spelling achievement, but that early diagnosis and individualized help will produce better habits of studying words.

Stegeman's (*43*) study of the importance of forgetting in spelling also does not argue for more instruction of the same kind but for properly spaced review.

The fact that curricula are based not only on social demands and on the characteristics of children, but also on the nature of the materials to be learned has been exemplified once again in Hanna and Moore's (*22*) suggestions for emphasizing a neglected phase of word study. They explored the possibilities of using the phonetic elements of our language in building a spelling program. More specifically, they recorded the consistency with which a certain speech sound is represented by one letter or combination of letters. They found that about four-fifths of the phonemes (speech sounds) in an elementary spelling vocabulary are represented by a regular spelling but that vowel phonemes are much less consistent in their spelling than initial consonant phonemes. They suggested that the spelling program be planned to present words close together in which the phonemes have a high degree of spelling consistency, thus encouraging the children to generalize about their spelling. They believed: "Our first job in teaching children to spell is to make certain that they can hear the sounds of the words they are to spell." (p. 336) Somewhat in opposition to these suggestions, Horn (*27*) found earlier that presenting spelling words in syllabified form offered no advantages.

As suggested above, in planning a spelling program factors of usage and need must be considered in addition to such auditory perception of word parts. However, further analysis of auditory abilities was one phase of the research reported in the next sections.

OTHER CHARACTERISTICS OF GOOD AND POOR SPELLERS

The first study of the characteristics of good and poor spellers by the present writer (*36*) compared two groups of good and poor spellers, matched for sex, mental ability and chronological age, but differing in spelling ability, on eighteen tests of different abilities. Pupils were given audiometer tests and the Betts perception series using the Keystone telebinocular. In general, the results indicated that the good spellers reliably exceeded the poor spellers on such factors as word pronunciation, reading comprehension and speed, reading accuracy and a systematic attack on new words to be learned. Poor spellers reliably exceeded good spellers in mispronunciations, and use of a letter-by-letter method in studying words.

The present study went beyond the first investigation in comparing good and poor spellers on auditory discrimination abilities, more specific visual perception, vocabulary and certain mental abilities included by the Thurstones in their *SRA Primary Mental Abilities* test.

During the last fifteen years knowledge of children's auditory abilities has been extended by a series of studies by Durrell and his students (*12*). Many of these have dealt with such abilities in relation to beginning reading, but Russell (*37*) has shown that there may be considerable overlap in verbal abilities basic to both reading and spelling in the primary grades. A recent study by Caffrey (*8*) of what he calls 'auding' has added considerably to what is known about auditory or listening skills. Caffrey experimented with tests of seven types of listening skills which included tests of vocabulary and several kinds of comprehension such as main idea and following directions. He also devised a test of auditory discrimination used in the present study. This test measured auditory abilities more specifically than those used in earlier investigations by Bond (*7*) and by Russell (*36*). It included three subtests: (1) distinguishing between pairs as the same or different such as *making making* and *shown sown*, (2) recognizing similar or different vowel sounds in words, as in *wine fight* and *move love*, (3) telling whether words were different in their beginning, middle or final sounds, as in *butter buzzer* and *pillow billow*.

The present investigation also supplemented the earlier one in extending measures of visual discrimination and vocabulary. Each child was given an experimental form of a visual perception test developed by the writer which, in a limited time, required him to mark words as the same or different. Sample items were:

here S D herd

portion S D portion

In addition, each child was given a series of seven vocabulary tests constructed by the writer in varied forms and measuring knowledge of words in different areas such as social studies, science, mathematics, sports and hobbies (*39*). Sample items were:

a. The word *copyright* usually appears in connection with (a) kitchen gadgets (b) silver coins (c) machines (d) books (e) income tax forms.

b. An *ounce* is commonly used to measure (a) potatoes

TABLE 1.—DIFFERENCES BETWEEN GOOD SPELLERS (UPPER 27 PER CENT) AND POOR SPELLERS (LOWER 27 PER CENT) OF APPROXIMATELY 250 CHILDREN IN THE FIFTH AND SIXTH GRADES

Measure	Mean Upper	Mean Lower	Difference	σD	t	$r_{x.sp}$
Auditory	43.82	27.76	16.06	1.96	8.19	.54
Visual	36.88	27.09	9.79	.91	10.76	.50
A + V	80.70	54.85	25.85	2.02	12.80	.66
California Reading						
Comprehension	71.20	41.58	29.62	6.17	4.80	.45
Vocabulary	72.06	23.47	48.59	4.48	10.84	.60
Experimental Vocabulary Tests						
Math	41.24	29.79	11.45	3.72	3.08	.42
Misc 1	74.66	35.62	39.04	4.75	8.22	.31
Misc 2	41.10	25.76	15.34	3.04	5.05	.48
Misc T	115.00	62.88	52.12	6.91	7.54	.66
Social Studies 1	30.57	18.05	12.52	2.54	4.93	.49
Social Studies 2	83.76	62.67	21.09	5.06	4.17	.42
Primary Mental Abilities						
Perception	62.11	37.05	25.06	4.86	5.16	.29
Reasoning	63.83	30.89	32.94	6.03	5.46	.51
Space	66.17	63.26	2.91	4.68	.62	.14
Total	60.81	27.05	33.75	5.96	5.66	.58

t = Fisher's t = difference between means divided by the standard error of that difference, sample size being taken into account in the computation equation.

$r_{x.sp}$ = Pearson r = correlation between spelling grade and the measure shown.

(b) baby food (c) fish (d) lumber (e) electricity (followed by two other harder items on the same word).

Other factors on which the groups were compared were the *California Reading Test* and the tests of the *SRA Primary Mental Abilities Test*. Table 1 gives a comparison of scores of approximately two hundred and fifty children in Grades V and VI who were in the top and bottom twenty-seven per cent of the distribution of scores on the spelling test of the *Progressive Achievement Test*. The results indicate that the group of good spellers exceed the group of poor spellers in the same grades on fourteen out of fifteen of the measures used, the t score indicating that the differences are significant at the .01 level. Only on the Space test of the Primary Mental Abilities Test is there no significant difference between the good and poor spellers. (This is a test of recognition of letter-like and geometric figures which have been rotated or otherwise shifted in position.) The results indicate that good spellers at the fifth- and sixth-grade levels tend to have superior auditory and visual perception, that they score higher in reading comprehension and a wide variety of vocabulary tests, and that they are superior in perception, reasoning and the total scores of the Primary Mental Abilities Test. Although this superior spelling ability is associated with superior mental ability as measured by the Thurstone test, a number of studies have shown that the correlation between spelling ability and general mental ability is not high, usually ranging between .20 and .55 (*36*). Even the relatively high correlation obtained in the present study (.58) suggests that factors of perception and word meaning as well as general mental ability contribute to spelling ability.

The relationships between the different variables and spelling ability are also indicated in the last column of Table 1. It may be noted here that the highest correlations with spelling scores are those of the combined auditory-visual test and a total miscellaneous vocabulary score.

Further analysis of the relation of the auditory and visual perception abilities to spelling was of interest, so scattergrams showing the relationships were constructed. The eta test of curvilinearity of regression (*33*) was applied and eta was found to differ significantly from the Pearson r in predicting spelling from auditory plus visual (A + V) score or auditory and visual scores alone. The results suggested that the Pearson r's of Table 1 somewhat underestimated the relationships and also that there is a somewhat lower correlation between spelling ability and auditory discrimination than there is between spelling ability and visual discrimination. Further, the results suggested that the relationship between spelling score and A + V score decreases as spelling score increases. Since the mean spelling score of the lowest twenty-seven per cent was 3.9 (in terms of grade) and of the highest twenty-seven per cent 7.8, it may be stated that spelling ability was more closely related to auditory and visual abilities around the third- and fourth-grade levels of ability than around the seventh- and eighth-grade levels of ability. Put another way, poor spelling ability in the fifth and sixth grades is closely related to poor discrimination of auditory and visual differences but high spelling ability at these levels is not closely related to superior discrimination. It may be that for children spelling at the average seventh- or eighth-grade level of achievement a number of factors other than auditory and visual discrimination abilities affect spelling ability.

The Pearson r between the auditory and visual tests was only .30, but the eta was .45. The regression lines for both variables were curvilinear but, in this case, predicting auditory from visual scores, or vice versa, was safer at the higher limits for both scores. That is, if a pupil scored high on the visual test he was likely to score high on the auditory test but if he scored low on either, prediction was less certain. The reliabilities of these experimental tests according to the Kuder Richardson formula 21 (*16*) were as follows:

Auditory	.93	Auditory + Visual	.93
Visual	.88		

CONCLUSION

During the last fifteen years considerable progress has been made in determining the words most useful to children and in discovering why some words are learned easily and others are harder to spell. Studies of eye-movements have indicated in more detail how children look at words in learning to spell them. Relationships between spelling and other language arts have been explored but other studies of methods of learning and instruction have been meager.

The present report added to a previous study of characteristics of good and poor spellers by further investigation of visual and auditory abilities in relation to spelling ability. It found that the upper twenty-seven per cent of approximately two hundred and fifty children in the fifth and sixth grades in terms of their spelling achievement significantly exceeded the lower twenty-seven per cent of the group on all fifteen tests used except the Space test of the SRA Primary Mental Abilities test. It discovered correlations between spelling ability and most of the other abilities tested, except Perception and Space on the Thurstone Test, which ranged in the .40's to .60's. It also found that visual and auditory discrimination abilities seem to be closely related to spelling ability around the third- and fourth-grade levels but not so closely related at the seventh- and eighth-grade levels of spelling ability.

REFERENCES

1) A. S. Artley, "Principles applying to the improvement of spelling ability," *Elementary School Journal*, November, 1948, 137–48.

2) — — —, "Research concerning interrelationships among the language arts," *Elementary English*, December, 1950, 527–37.

3) L. P. Ayres, *A Measuring Scale for Ability in Spelling*. New York: Russell Sage Foundation, 1915.

4) E. A. Betts, *Spelling Vocabulary Study: Grade Placement of Words in Seventeen Spellers*. New York: American Book Co., 1940.

5) — — —, "Inter-relationships of reading and spelling," *Elementary English Review*, January, 1945, 13–23.

6) — — —, *A Second Spelling Vocabulary Study: Grade Placement of Words in Eight Recent Spellers*. New York: American Book Co., 1949.

7) G. L. Bond, *The Auditory and Speech Characteristics of Poor Readers*. Contributions to Education No. 657. New York: Bureau of Publications, Teachers College, Columbia University, 1935.

8) J. G. Caffrey, *Auding Ability as a Function of Certain Psychometric Variables*. Doctor's Dissertation, University of California, Berkeley, 1953.

9) H. A. Curtis, and E. W. Dolch, "Do spelling books teach spelling?" *Elementary School Journal*, April, 1939, 584–92.

10) E. Dale, *Bibliography of Vocabulary Studies*. Columbus: Bureau of Educational Research, Ohio State University, 1949.

11) E. W. Dolch, *Better Spelling*. Champaign, Illinois: Garrard Press, 1942.

12) D. D. Durrell and H. A. Murphy, "The auditory discrimination factor in reading readiness and reading disability," *Education*, May, 1953, 556–60.

13) J. A. Fitzgerald, *A Basic Life Spelling Vocabulary*. Milwaukee: Bruce Publishing Co., 1951.

14) — — —, *Teaching of Spelling*. Milwaukee: Bruce Publishing Co., 1951.

15) ———, "Teaching of spelling," *Elementary English*. February, 1953, 79–85.

16) G. J. Froehlich, "A simple index of test reliability," *Journal of Educational Psychology*, May, 1941, 381–85.

17) A. I. Gates, *A List of Spelling Difficulties in 3876 Words*. New York: Bureau of Publications, Teachers College, Columbia University, 1937.

18) — — —, "Reading in relation to spelling," *Texas Outlook*. January, 1946, 8–10.

19) L. C. Gilbert, *An Experimental Investigation of Eye Movements in Learning to Spell Words*, Psychological Monograph, No. 3, Princeton, N. J.: Psychological Review Co., 1932.

20) L. C. Gilbert, and D. W. Gilbert, *Training for Speed and Accuracy of Visual Perception in Learning to Spell Words*. University of California Publications in Education, Vol. VII, No. 5, Berkeley: University of California Press, 1942.

21) W. S. Guiler and G. A. Lease, "An experimental study of methods of instruction in spelling," *Elementary School Journal*, December, 1942, 234–38.

22) P. R. Hanna and J. T. Moore, "Spelling, from spoken word to written symbol," *Elementary School Journal*, February, 1953, 329–37.

23) G. H. Hildreth, "Interrelationships among the language arts," *Elementary School Journal*, June, 1948, 538–49.

24) — — —, "An evaluation of spelling word lists and vocabulary studies," *Elementary School Journal*, January, 1951, 254–65.

25) L. S. Hollingworth, *The Psychology of Special Disability in Spelling.* Contributions to Education No. 88. New York: Teachers College, Columbia University, 1918.

26) E. Horn, "Spelling," *Encyclopedia of Educational Research*, p. 1247–64. New York: Macmillan Co., 1950.

27) T. D. Horn, *The Effect of a Syllabic Presentation of Words Upon Learning to Spell.* Doctor's Dissertation, State University of Iowa, 1947.

28) — — —, "That Straw Man: The Spelling List," *Elementary English*, May, 1952, 265–67.

29) D. M. Knoell, and C. W. Harris, "Factor analysis of spelling ability," *Journal of Educational Research*, October, 1952, 95–111.

30) G. C. Kyte, "A core vocabulary for the primary grades," *Elementary School Journal*, November, 1943, 157–66.

31) G. C. Kyte, and V. M. Neel, "A core vocabulary of spelling words," *Elementary School Journal*, September, 1953, 29–34.

32) I. Lorge, *The Semantic Count of the 570 Commonest English Words.* New York: Bureau of Publications, Teachers College, Columbia University, 1949.

33) Q. McNemar, *Psychological Statistics.* New York: John Wiley and Sons, 1949.

34) A. M. Nichols, "The analysis and correction of spelling difficulties," *Elementary School Journal*, November, 1949, 154–61.

35) H. D. Rinsland, *A Basic Vocabulary of Elementary School Children.* New York: Macmillan Co., 1945.

36) D. H. Russell, *Characteristics of Good and Poor Spellers.* Contributions to Education No. 727. New York: Bureau of Publications, Teachers College, Columbia University, 1937.

37) — — —, "A diagnostic study of spelling readiness," *Journal of Educational Research*, December, 1943, 276–83.

38) — — —, "Spelling ability in relation to reading and vocabulary achievement," *Elementary English Review*, January, 1946, 32–37.

39) — — —, *The Dimensions of Children's Meaning Vocabularies in Grades Four through Twelve.* University of California Series in Education. Berkeley: University of California Press, 1954.

40) G. Spache, "Critical analysis of various methods of classifying spelling errors," *Journal of Educational Psychology*, February, 1940, 111–34.

41) J. L. Sparrow, *A Study of the Relationship between Spelling Ability and Reading Ability.* Doctor's Dissertation, State University of Iowa, 1951.

42) R. G. Stauffer, "Study of prefixes in the Thorndike List to establish a list of prefixes that should be taught in elementary schools," *Journal of Educational Research*, February, 1942, 453–58.

43) W. H. Stegeman, *Measurement and Control of Forgetting Rates in Spelling.* Doctor's Dissertation, University of California, Berkeley, 1947.

44) C. R. Stone, *Stone's Graded Vocabulary for Primary Reading.* St. Louis: Webster Publishing Co., 1941.

45) E. L. Thorndike, *Teaching of English Suffixes.* Contributions to

Education No. 847. New York: Bureau of Publications, Teachers College, Columbia University, 1941.

46) E. L. Thorndike and I. Lorge, *The Teacher's Work Book of 30,000 words*. New York: Bureau of Publications, Teachers College, Columbia University, 1944.

47) A. Townsend, "Investigation of certain relationships of spelling with reading and academic aptitude," *Journal of Educational Research*, February, 1947, 465–71.

48) A. G. Wesman and H. G. Seashore, "Frequency versus complexity of words in verbal measurement," *Journal of Educational Psychology*, November, 1949, 395–404.

49) D. E. Wilson, *The Development of Spelling in the Elementary School Curriculum*. Doctor's dissertation, University of California, Los Angeles, 1951.

50) P. K. Wilson, "Comparison of words used in spelling texts with those used by children," *California Journal of Educational Research*, March, 1953, 84–86.

51) Sister M. R. Wolff, "A study of spelling errors with implications concerning pertinent teaching methods," *Elementary School Journal*, April, 1952, 458–66.

52) G. A. Yoakam, "The Relation of Reading to Spelling and Handwriting," *Report of the Sixth Annual Conference on Reading*, p. 72–80. Pittsburgh: University of Pittsburgh Press, 1950.

THE ORGANISMIC AGE CONCEPT

PAUL BLOMMERS, LOTUS M. KNIEF, AND J. B. STROUD

State University of Iowa

The Organismic Age concept has had considerable appeal to elementary teachers. Perhaps one reason for this is the fact that its proponents have put it forward in a setting of enlightened school practices. In this paper we are not concerned with these practices but with the validity of the OA concept.

Long since accepted is the idea that a child's readiness for reading, arithmetic, third-grade science, and so on is a function of his mental development, total mental development. In this, the MA is the most important single component. But social development, emotional maturity, before-school experiences, language development, motor coördination, and perhaps other factors are also important. The OA is an average of age scores of mental data and age scores of anatomical and physiological data, which hereafter we shall call physical data. Commonly, OA is the average of mental age, reading age, dental age, grip age, metacarpal age, vital capacity age, height age, and weight age, although there is no hard and fast set of component measures. In a sense it combines measures of what is commonly called mental growth and measures of what we may call physical growth; and, because of the disproportionate number of measures of the latter kind, the average, OA, is weighted heavily in that direction. It is with the feasibility in school practice of making prediction of achievement, determining readiness, appraising achievement of pupils in terms of their anatomical and physiological development, as well as in terms of measures of mental development that we are concerned. Can we make better predictions or interpretations (which amounts to the same thing) by so doing?

Obviously the first questions one would ask in this connection are: Are measures of physical growth related (1) to mental growth, and (2) to academic achievement? Moreover, if found to be related, are they related to an extent that makes any practical difference?

They are, of course, related in a spurious way. All the measures that go into OA are related one to another in this way. Any growth process requires time. Childhood is the period of growth. Most of the growth functions begin in early childhood, even before birth, and continue until maturity is reached. Naturally there is a certain relatedness or togetherness in the various growth functions. Children who wear Number 5 shoes are taller, heavier, have bigger bones, stronger grip, and more teeth than children who wear Number 2 shoes. They also can read better, know more about world geography, and have higher MA's. Educational development and mental development take time also. This is the meaning of growth. All the time the child is getting older. He does not grow because he gets older; but he must get older before he can grow to a larger size. Childhood is not only the time for physical growth; it is also the time for mental growth, and growth in knowledge of school subjects. If, in longitudinal studies, we plot curves showing growth, or yearly increments, over a number of years, we would necessarily find a relatedness in all the functions measured because elapsed time is common to all the measures.

But children do not all grow at the same rate. All do not reach the same point at age seven, for example. At age seven, some have MA's of 8, 9, and 10; others have MA's of only 6, 5, or 4. Some are as tall as the average eight-, nine-, or ten-year-old; others are no taller than the average six-, five-, or four-year-old. They likewise vary in all other aspects of physical growth and in academic achievement. Everywhere we know that academic achievement is related to mental growth. If we pick out the brightest fourth and dullest fourth of our seven-year-old group, we will find that we also have the highest and lowest fourths in achievement, and that the differences in achievement will be large. This is what we mean by saying that mental development and educational development are related one to the other. We can predict this in advance for any normal group of seven-year-old pupils. This, again, is what we mean by related.

If physical growth is to be of any help, those same seven-year-old pupils who are advanced in educational development and mental development must turn out to be the tallest, heaviest, have the most teeth, the best-developed bones, the strongest grip, and so on. Those who are retarded in education and mental development must show underdevelopment in these physical traits and in significant amounts. If this does happen, we can say physical development is related to educational and mental development, and we can demonstrate the presence and amount of relationship by statistics. If we cannot demonstrate the relationship statistically, we have no right to say they are related.

Research on this important problem dates back to around 1900. All the measures of physical growth that have been here mentioned have been correlated with measures of mental and educational growth—most of them several times—by various investigators. These investigations have most generally yielded correlation coefficients at or near zero, when CA is held constant. By contrast, investigators everywhere have found substantial and significant relationships between measures of mental growth and educational development. Both of these generalizations are so well known and commonly agreed upon, it seems unnecessary to treat this subject historically or to cite extensively the experimental investigations (*1*). Height age and weight age, and they alone of the physical measures, have been found to correlate significantly with MA, CA held constant. The average of the obtained coefficients is something on the order of .20. Predictions made on correlation coefficients of this magnitude are not more than two per cent better than chance.

Dearborn and Rothney report coefficients between intelligence test scores and various measures of physical growth for five hundred and thirty-three sixteen-year-old boys as follows (*2*):

	Standing Height	Weight	Iliac	Chest Depth	Chest Width
Intelligence	.224	.137	.078	.060	.138

The multiple correlation between intelligence test scores and the five physical measures was found to be R = .247.

The writers have access to some data collected for another purpose by Choitz (*3*). These include Otis MA's, reading scores, arithmetic scores, height age scores, weight age scores, and dental age scores, among others, for a sample of pupils in grades 4, 5, and 6—about forty pupils in each grade. By a within-grade design, which tends to hold chronological age more or less constant, intercorrelations and multiple correlations were computed. Multiple correlations were obtained as follows:

	MA, alone	MA, HA	MA, HA, WA	MA, HA, WA, DA
Reading	.6451	.6508	.6543	.6592
Arithmetic	.5507	.5520	.5559	.5560

The total effect of the addition of the three physical measures, HA, WA, and DA raised the correlation coefficient between reading achievement and MA and arithmetic achievement and MA by approximately .01 (*4*). It is likely that the addition of GA, CA, and VcA would have raised the multiple correlation by perhaps a like amount. A change of this amount in a coefficient has neither statistical nor practical significance. These findings are in agreement with those obtained by Gates many years ago (*5*). He obtained a correlation of .60 between MA and educational achievement scores. A multiple correlation procedure involving, in addition to MA, bone ossification ratio, height, weight, chest girth, lung capacity, grip, and nutritional status, yielded a coefficient of .63, an increase of .03.

Proponents of OA sometimes write in such a vein as to suggest a belief in some kind of general urge to grow, or some kind of general growth force, which all growth processes must obey. It appears as if a child who has only eighty per cent of a normal growth force will grow only eighty per cent as fast as a child with a normal amount of this element; or, one with one hundred and twenty per cent of normal will grow at a rate of one hundred and twenty per cent. Obviously there is no single force which both mental and physical growth factors must obey and obey alike, because as we have seen, the rates of the two kinds of growth are essentially unrelated.

There is some relatedness in rate of growth among various physical measures; and also among the various traits that we speak of as making up intelligence. Also, with respect to each set of measures, growth tends to be regular and continuous; tends to proceed at the rate at which it starts out. Correlation coefficients of .80 and .90 have been obtained between measures of height taken several years apart, say at age seven and at age seventeen (*6*). Coefficients of .80 to .90 have also been reported between scores on intelligence tests obtained ten years apart (*7*, *8*).

Rate of growth, physical and mental, is probably governed both by hereditary and environmental factors. Heredity is set in the young infant, actually at conception, and remains constant. While environmental factors are, of course, subject to change, they probably remain fairly constant for most children. These two factors should give stability and regularity to growth. A child who is going to be large when he grows up, is large relative to others of his age all along—at six, twelve, or eighteen. Similarly, one who is going to be small at maturity is small at all stages of growth. This is the meaning of the positive correlations obtained between growth measures taken years apart. A child who is going to be bright or dull at maturity is correspondingly bright or dull at six, twelve, or eighteen.

Among physical traits there is no compelling necessity for a large or small child to be large or small all over. This is reflected in the fact that we think it desirable to utilize several different measures of physical growth. However, over-all, the various measures tend to be positively intercorrelated. The writers obtained (with Choitz's data) the following intercorrelations: HA and WA, .57; HA and DA, .24; WA and DA, .10. Dearborn and Rothney obtained an average intercorrelation of .57 among Standing Height, Weight, Iliac, Chest Depth, and Chest Width. Intercorrelations among most of the measures that are commonly associated with OA appear to be considerably lower than this average. Lowell and Woodrow reported a coefficient of .20 between carpal age and number of permanent teeth (*9*). Gates obtained coefficients between carpal scores and lung capacity of .30, and strength of grip of .25; between strength of grip and height, .45, weight, .40, lung capacity, .46. Thus as Paterson observed, "Various aspects of physical growth are shown to be far from unitary."

We see that the unity of growth, one of the tenets of OA theory, is a somewhat questionable concept when applied to physical growth alone. It has very little meaning with respect to 'total growth'. Proponents of OA theory have stated that the various aspects of growth may be viewed as interchangeable samples of total growth. It is further stated that " . . . . the various attributes in an individual tend to cluster about a center of gravity of growth of that individual and that the freedom to vary is restricted (*10*)."

In exposition of this kind a writer is not ordinarily concerned with expressing a view as a testable hypothesis. When one interested in making a test tries to restate it in such a way that it may be tested statistically, one runs the risk of making a statement that is at variance with the author's intended meaning. The statement just quoted seems to say that the various aspects of growth are positively interrelated and suggests the interpretation that any sample of K aspects of growth may be substituted for any other such aspects in any prediction based on such measures as the one intended. This interpretation may not be precisely the one intended, but it must be admitted that it is consistent with and follows naturally from the general OA concept. We have already examined this interpretation in the light of evidence yielded by correlation studies. We shall now present findings derived from the application of a somewhat different type of statistical technique.

This technique as we applied it involved an analysis of the total variability among three selected growth measures for the pupils in each of three grades, again using Choitz's data. The three growth measures so employed were chosen to represent each of the three major types of growth measures included in the OA index. One was a measure of mental growth, Mental Age; one was a measure of physiological growth, Weight Age; and one was a measure of anatomical growth, Dental Age. These measures were admittedly selected because of their known low intercorrelations. However, this choice is not in any way inconsistent with the claim that the various aspects of growth are interchangeable samples of total growth.

It may be argued that the use of only three growth measures, as compared with the greater number commonly utilized, does not provide a valid test of the general hypothesis or 'view' in question, i.e., that K should be greater than three. To this it can be said that one method of testing a hypothesis is to determine whether or not there exist any specific exceptions to it. Failure to discover such exceptions does not prove the hypothesis but, as possible exceptions are eliminated, confidence in the hypothesis develops. On the other hand, the discovery of even a single exception does discredit the hypothesis, at least to the extent of demanding its revision to allow for this exception.

It was with these principles in mind that analysis was made of MA, WA, and DA scores which would test the tenet of 'unity of growth', i.e., that the various attributes in a given individual tend to cluster about a center of gravity with a concomitant restriction of variation.

The technique employed is that of component analysis, which has been developed in connection with the technique of analysis of variance. In a general way, this technique analyzes the total variation among all the growth measures available for all the individuals into component parts, and provides an estimate of the proportions of this total variation which may be ascribed to each part. The total variation was analyzed into two components; namely, one, a component reflecting the variation from individual to individual among the means of the three growth measures, and the other a component reflecting the variations among the three growth measures within the respective individuals. In this case the total variation referred to was determined separately for each of the grade levels—fourth, fifth, and sixth.

It will be noted that the first component cited, in effect, deals with variations in OA from child to child, insofar as OA can be estimated from a pooling of three samples of it. The second component reflects the variation among an individual child's growth measures. It, in effect, represents an averaging of such variations obtained separately for each child. Or, in the language of the proponents of OA theory, the first component represents variation from center of gravity to center of gravity of the different children. The second component represents a sort of average of the variations among the various scores which comprise each child's cluster.

If the view under test is to prevail, it necessarily follows that the first of these components must account for a significant and substantial proportion of the total variation among the growth measures. The per cents associated with each grade group are as follows:

Grade	1st Component	2nd Component
4	9%	91%
5	12%	88%
6	15%	85%

Thus the proportion of total variation accounted for by the first component, among individuals, is far from substantial. Since so large a proportion of the total variation is accounted for by the second component, within individuals, it would appear that, in this case, the growth measures are not interchangeable samples of total growth and that freedom to vary is not restricted—certainly not restricted much.[1]

Moreover, if we may assume that the MA, WA, and DA scores for a child constitute a random sample from a normally distributed population or complete cluster of such growth scores,[2] then it is possible to test the statistical significance of the differences among the means from child to child. These differences were not found to be significant at any of the three grade levels studied. In other words, differences among children's OA scores may be explained in terms of chance or accidental variations among the individual growth scores of the children involved. If this is the case, then the true centers of gravity based on theoretically complete clusters of scores will not differ from child to child within a school grade group and hence, OA scores which are but estimates of these true centers of gravity do not constitute a meaningful basis for discriminating one child from another.

---

[1] It will be noted, inasmuch as the same three age scores were determined for each child, that a component analysis involving the proportions of total variation which may be ascribed to (1) variation in OA from child to child, (2) variations in mean age from attribute to attribute and (3) interaction between OA and attribute could have been employed. It was felt, however, that the two-component analysis selected represented a model more consistent with the general 'views' under examination than the three-component analysis here mentioned.

Since it is obvious, further, that the children in a grade group do differ with respect to single growth characteristics, the fact that these same children do not differ significantly in terms of the average of a number of growth characteristics, is evidence of nothing more than the chance association of individual growth measures within an individual child. There is no systematic tendency, in other words, for a child advanced in Mental Age to be advanced also in Weight Age or Dental Age.

REFERENCES

1) D. G. Paterson, *Physique and Intellect*. New York: Appleton-Century-Crofts, 1930.

2) W. F. Dearborn and J. W. M. Rothney, *Predicting the Child's Development*. Cambridge (Mass.): Sci-Art Publishers, 1941.

3) John F. Choitz, *A Survey for Handicaps Among Approximately 400 Elementary School Children in Johnson County, Doctor's Dissertation*, State University of Iowa, 1952.

4) Lotus M. Knief, *A Study of the Relationship between Achievement in Reading and Arithmetic and Various Growth Factors, Master's Thesis*, State University of Iowa, 1953.

5) Arthur I. Gates, "The nature and educational significance of physical status and of mental, physiological, social and emotional maturity," *Journal of Educational Psychology*, Vol. 15, 1924, pp. 329–358.

6) H. V. Meredith, "The prediction of stature of north European males through the elementary school years," *Human Biology*, 1936, Vol. 8, pp. 279–283.

7) H. E. Jones, "Environmental Influences on Mental Development," *Manual of Child Psychology* (L. Carmichael, Ed.), pp. 582–632.

---

[2] The writers are fully aware of the questionable aspects of this assumption. The fact remains, however, that it is not inconsistent with the 'view' of OA theory under examination.

8) N. Bayley, "Some increasing parent-child similarities during the growth of children," *Journal of Educational Psychology*, 1954, Vol. 45, pp. 1–21.

9) F. Lowell and H. Woodrow, "Some data on anatomical age and its relation to intelligence," *Pedigogical Seminary*, 1922, Vol. 19, pp. 1–15.

10) W. Olson and Byron O. Huges, "Concepts of growth: their significance to teachers," *Childhood Education*, 1944–45, Vol. 21, p. 55.

READING IMPROVEMENT AND COLLEGE GRADES: A FOLLOW-UP

DONALD E. P. SMITH AND ROGER L. WOOD

University of Michigan

Evaluation of corrective reading treatment, like other therapies, is hampered by the requirement of humane treatment of subjects. Rigorous experimental design is therefore difficult to maintain. Nevertheless, criticisms (*8*, *13*) of current evaluative studies are justified. Since assignment of subjects to experimental and control groups is based on convenience (*5*) or self-determination (*14*), major selective factors, e.g., motivation of subjects in each group, are often left uncontrolled. (The citations are merely representative; a number of others might have been used.) Furthermore, changes in test performance are reported as changes in ability, a questionable assumption.

Further criticisms of current assessment studies relate to the utilitarian aim of effect on academic achievement (*8*) and permanence of performance gains, i.e., the 'hothouse' effect (*1*, p. 74). While slight differences in grade point average greater than those of control groups have been reported recently (*2*, *4*, *5*, *14*), an overwhelming number of writers in the past either have not reported grade changes or have found little or no increase in academic status concomitant with increase in reading performance (*3*). Whether failure to report was due to oversight, lack of data, or absence of positive results is unknown.

The present study was designed with the above criticisms in mind. An attempt was made to determine changes in reading performance, over a brief and over an extended interval, and in academic status of students who volunteered or were referred for a corrective reading course at a state university.

PROCEDURE

Subjects of the study were seventy-four male and female freshmen of the University of Michigan who came voluntarily or by referral to the Reading Improvement Division of the Bureau of Psychological Services during its first semester of operation, and who completed the course.[1] Another twenty-one freshmen who attempted to register could not be enrolled due to lack of facilities. This group, equivalent to the experimental subjects in *ACE Psychological Examination for College Freshmen* mean scores, constituted a motivated, non-treatment control group. A third group comprised a representative sample of freshmen. This group was assembled by choosing every fiftieth name from an alphabetical list of freshmen.

A further sample of thirty was drawn randomly from the experimental group of seventy-four for follow-up testing to determine permanence of gain in reading performance sixty weeks after termination of the course.

The training procedures in the reading laboratory were similar to those used in most collegiate programs. Students met in small groups for two one-hour sessions each week for ten weeks and were encouraged to work by themselves in the laboratory for two one-half-hour periods. There was considerable variation in the extent to which this suggestion was followed. One of the weekly hour meetings occurred in a seminar room and was devoted to instruction in the psychology of reading, e.g., methods for attaining rate flexibility, in study techniques, e.g., utilization of Francis P. Robinson's *Survey Q3R* method (*7*), and in preparing for and writing essay and short answer examinations. Discussion of common problems was usual, and each session included mimeographed exercises concerning vocabulary, comprehension skills, critical reading or the like. It was concluded with a timed reading from Ruth Strang's *Study Type of Reading Exercises* (*12*) or from Wilking and Webster's *A College Developmental Reading Manual* (*15*). The second hour was spent in supervised laboratory practice with mimeographed work sheets, fiction and non-fiction books, textbooks, pacing devices and a tachistoscope. The skills and principles have been described elsewhere (*10*).

Assessment of changes in reading performance and in academic status concomitant with or following treatment was determined as follows:

1) Reading Performance: *Coöperative Reading Test C2* (Vocabu-

---

[1] Attrition was less than ten per cent. While such a loss might distort the results, it appears from present evidence of the character of drop-outs that such distortion would depress rather than inflate results. Those who drop out tend to be 'fast improvers' (*9*).

TABLE I.—CHANGES IN MEAN READING PERFORMANCE OF COLLEGE STUDENTS AFTER INTERVALS OF TRAINING (Ten Weeks) AND NO TRAINING (Sixty Weeks)

Skill	Coöperative Reading Test C2 (N = 27)								
	Initial (Form T) Scaled Score	σ	%ile†	Final (Form Y) Scaled Score	σ	%ile†	Follow-up (Form T) Scaled Score	σ	%ile†
Vocabulary	54.3	7.1	38	55.9	6.9	45	56.5	7.6	40
Level	55.9	6.6	49	58.9	6.0	55	61.4*	7.1	60
Speed of Comprehension	56.5	8.2	40	62.6*	11.3	62	64.8**	10.6	69

* Increase over initial mean significant at .05

** Increase over initial mean significant at .01.

† Percentile equivalents are based upon test manual norms as follows: Initial and Final: "entering freshmen;" Follow-up: "freshmen (end of year)"

lary, Level of Comprehension, Speed of Comprehension). *Traxler High School Reading Test*, Part I, (Rate, Comprehension).

2) Academic Status: Registrar's record of grade-point average (GPA). Probationary status. Forced withdrawal.

A final category, voluntary withdrawal, includes those subjects who, at the time of transfer or withdrawal, were maintaining an acceptable grade point average of 2.0 or above.

RESULTS

Changes in reading status and permanence of gains are reported in Tables I and II and are presented graphically in Figure 1. Table I indicates the *Coöperative Reading Test C2* mean scaled scores (standard scores by which equivalence among forms is attained) of experimental subjects at the initial (pre-training), final (post-training) and follow-up (lapse of 60 weeks) testings. Of the thirty subjects drawn at random from the original seventy-four[2], two had voluntarily withdrawn from the University at the end of their third semester and one declined to coöperate. Changes in reading performance were negligible for Vocabulary while Level of Comprehension differed significantly (.05) between Initial and Follow-up testing. Speed of Comprehension, the skill most sensitive to the

---

[2] No differences approaching significance were found between the sample of thirty and the total group of seventy-four in reading scores.

TABLE II.—CHANGES IN READING PERFORMANCE OF COLLEGE STUDENTS AFTER INTERVALS OF TRAINING (TEN WEEKS) AND NO TRAINING (SIXTY WEEKS)

Traxler High School Reading Test

(N = 27)

Skill	Initial (Form A)	σ^*	Final (Form B)	σ	Follow-up (Form A)	σ
Rate (wpm)	219.4	37.7	272.2**	55.9	297.2**	85.7
Comprehension (%)	75.6	10.2	70.2	12.2	76.3	12.1
Rate of Comprehension	165.9		191.1		226.8	

* Differencies in σ_{rate} for Initial-Final and Final-Follow-up are significant at .05.

** Increase over Initial Rate significant at .01.

training methods used, showed increases of 22 and 29 percentiles over ten weeks and sixty weeks, respectively. (For the total group, N=74, all increases were significant at or beyond .05 at the end of training.)

Table II includes similar results recorded on the *Traxler High School Reading Test*. An initial rate of 219.4 words per minute was found to be 272.2 ($P < .01$) and 297.2 ($P < .01$) at the final and follow-up testings. At the same time comprehension (immediate retention of story content) showed negligible change.[3] It is of interest to note the continued large increases in standard deviation ($P < .05$), a result conforming to the principle that one effect of formal education is an increase in individual differences. Rate of Comprehension in an index of change which minimizes the effect of an undue increase in rate at the expense of understanding. It provides a useful index of progress despite certain statistical limitations.

Figure 1 illustrates the same changes in performance during the training program and the permanence of gains. Vocabulary scores showed no change. Speed of Comprehension scores increased most rapidly over ten weeks and continued to increase as did Level of Comprehension scores and the Traxler Rate of Comprehension index. Since the *Coöperative Reading Test C2* results are presented in

---

[3] An independent study of forms A and B of the *Traxler High School Reading Test* casts doubt on their assumed equivalence when used with adults (*11*).

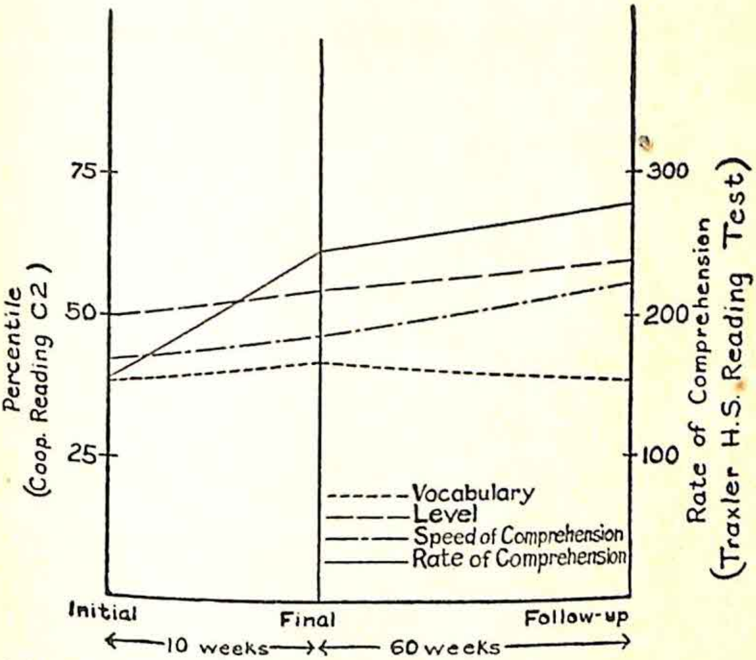


FIG. 1.—Reading Performance of Experimental Subjects after Ten Weeks of Training Followed by Sixty Weeks of No Training.

the figure as percentile equivalents, the differences reflect increases beyond those normally anticipated as a result of the academic experience alone, e.g., while a scaled score of 63 in Speed of Comprehension is equivalent to the 55th percentile for entering freshmen, a score of 64 is equivalent to the same percentile at the end of the year.

Changes in academic achievement are reported in Table III. While no significant change occurred from the first to the second semester in the grade point averages of the Control and Representative Freshmen groups, a significant increase ($P < .01$) occurred in the Experimental group. A GPA of 2.0 corresponds to a grade of C, 3.0 to a B, etc. It will be noted that the grade point averages of the second semester do not include the standing of students withdrawing from the University. Therefore, the results are probably conservative estimates of the true differences among the three groups.

While Experimental and Control differed by .11 ($P < .05$) in status at the end of the semester during which training occurred, the difference increased to .25 ($P < .01$) in favor of the Experimen

TABLE III.—COMPARATIVE MEAN ACADEMIC ACHIEVEMENT AMONG EXPERIMENTAL, CONTROL AND REPRESENTATIVE FRESHMEN GROUPS

Group	1st Semester		2nd Semester*		Diff.	Sig.
	N	G.P.A.	N	G.P.A.		
Experimental	74	2.25	71	2.41	+.16	$P < .01$
Control	21	2.14	17	2.16	+.02	NS
Freshmen	46	2.29	40	2.22	−.07	NS

	1st Semester		2nd Semester	
	G.P.A. Diff.	Sig.	G.P.A. Diff.	Sig.
Experimental v Control	.11	$P < .05$	.25	$P < .01$
Experimental v Freshmen	.04	NS	.19	$P < .01$

* Does not include subjects who withdrew from the University.

tal group by the end of the second semester. A similar increase in achievement of the Experimental group over the Representative Freshmen group is shown ($P < .01$).

Amount and kind of academic deficiency is reported in Table IV. Scholastic aptitude, inferred from group mean raw scores (converted to percentile equivalents) of the *ACE Scholastic Aptitude Test for College Freshmen*, was similar for the three groups. However, the Experimental group included a significantly smaller percentage on probation than did the Freshmen (10.8 per cent and 32.6 per cent, $P < .05$). The Experimental group also included a significantly smaller percentage in the combination category of probationary and forced withdrawal than did either the Control group (12.2 per cent and 33.3 per cent, $P < .05$) or the Freshmen group (12.2 per cent and 41.3 per cent, $P < .01$).

The superiority of the Control group over the Representative Freshmen, while not significant, suggests the commonly suspected difference in motivation, 'fear-induced drive', aspiration-level, or the like between those who apply to a reading service and those who do not apply.

TABLE IV.—COMPARATIVE MEAN ACADEMIC DEFICIENCY OF EXPERIMENTAL, CONTROL AND REPRESENTATIVE FRESHMEN GROUPS (2ND SEMESTER)

Group*	N	ACE %ile			% Below 2.0**	% Forced Withdrawal	% Total Deficiency	(% voluntary wd.)
		Q	L	T				
Experimental	74	56	40	46	10.8	1.4	12.2	(3.7)
Control	21	60	40	48	23.8	9.5	33.3	(9.5)
Freshmen	46	45	51	49	32.6	8.7	41.3	(4.3)

	Below 2.0		Total Deficiency	
	Diff.	Sig.	Diff.	Sig.
Experimental v Control	13.0%	NS	21.1%	$P < .05$
Experimental v Freshmen	21.8%	$P < .05$	29.1%	$P < .01$

* All groups were nearly identical at the mean in academic load.

** Does not include subjects categorized under 'forced withdrawal'.

DISCUSSION

Within the limits of sample size, subject characteristics and treatment methods, it may be concluded that changes did occur in reading performance and academic status concomitant with and following application of the corrective techniques used. It is the feeling of the writers that fulfillment of certain conditions may have been responsible for the results. First, it may be too much to expect that improved reading performance in a laboratory will result in improved grades. The appropriate skills to be developed seem to be those directly involved in the study and examination situations. Second, if we are to expect permanence of improvement in reading performance, it seems apparent that provision must be made for continued practice in attempting to read efficiently. The reader will recognize the two conditions as applications of familiar learning principles relating to transfer and performance.

While the above conditions may be necessary, they are probably far from sufficient to bring about the maximum contribution to students. It is our opinion that important factors hardly touched as yet are those of personality structure-learning interaction (*9*), perceptual development, treatment variables (especially the instructor as a variable) and diagnosis.

SUMMARY AND CONCLUSIONS

Evaluation of collegiate reading improvement services have been justifiably criticised for inadequate experimental design. The present study concerns changes in reading performance, permanence of gains and concomitant changes in academic status of clients of a university reading service. An attempt was made to control the main variables confounding the results of previous investigations.

Major conclusions are as follows:

1) Significant gains in performance (though not necessarily in skill) result for those aspects of reading which are emphasized in training.

2) Performance gains are maintained and, possibly, increased after a lapse of time (sixty weeks) with no formal training when continued practice is encouraged.

3) Significant superiority in academic status (increasing with time) is demonstrated by experimental subjects over both control and representative freshmen subjects when study and examination skills are emphasized during the training period.

REFERENCES

1) I. H. Anderson and Dearborn, W. F. *The Psychology of Teaching Reading*. New York: Ronald Press Co., 1952.

2) W. B. Barbe. "The effectiveness of work in remedial reading at the college level." *J. educ. Psych.* 43, 1952, 229–237.

3) R. W. Kilby. "The relations of a remedial reading program to scholastic success in college." *J. educ. Psych.*, 36, 1945, 513–534.

4) Dorothy J. McGinnis. "Corrective reading: A means of increasing scholastic attainment at the college level." *J. educ. Psych.*, 42, 1951, 166–173.

5) G. J. Mouly. "A study of the effects of a remedial reading program on academic grades at the college level." *J. educ. Psych.*, 43, 1952, 459–466.

6) L. C. Pressey and Pressey, S. L. "Training college freshmen to read" *J. educ. Res.*, 21, 1930, 203–211.

7) F. P. Robinson. *Effective Study*. New York: Harper and Bros., 1946.

8) H. A. Robinson. "A note on the evaluation of college remedial courses." *J. educ. Psych.*, 41, 1950, 83–96.

9) D. E. P. Smith and Wood, R. L. "Noncognitive personality dimensions as indices of reading improvement at the college level." (abs.) *Amer. Psych.*, 8, 1953, 438.

10) —, *Help Yourself to Efficient Reading*. Ann Arbor: University of Michigan Press, 1953.

11) —, "A Note on Equivalent Forms." (unpub. study, 1953).

12) Ruth Strang. *Study Type of Reading Exercises*. New York: Bureau of Publications, Teachers College, Columbia University, 1935.

13) Francis Triggs. "Current problems in remedial reading for college students." *School and Society*, 53, 1941, 376–379.

14) F. L. Westover. *Controlled Eye Movements versus Practice Exercises in Reading*. (Teach. Coll. Cont. to Educ., No. 917) New York: Bureau of Publications, Teachers College, Columbia University, 1946.

15) S. V. Wilking and Webster, R. G. *A College Developmental Reading Manual*. Boston: Houghton Mifflin Co., 1943.

PROBLEM-SOLVING BY TEAMS AND BY INDIVIDUALS IN A FIELD SETTING[1]

IRVING LORGE, JACOB TUCKMAN, LOUIS AIKMAN,
JOSEPH SPIEGEL AND GILDA MOSS

Teachers College, Columbia University

In a previous study, the relationship between the method of presentation of a field problem and the quality of the written solutions was investigated (*1*). The field problem employed, the Mined Road Problem (*2*), adapted from that developed by the Office of Strategic Services, requires the formulation of a plan of action for getting a group of five men across the road mined with supersensitive enemy mines which can neither be neutralized nor dug up. The road is twelve feet wide, bordered with trees about forty feet tall. Scattered about one side of the road is a variety of potentially usable material and debris, including beams, boards, ropes, discarded auto tire, pulley, etc. As adapted, the Mined Road was presented at four laboratory levels of remoteness from reality: (1) verbal description (2) photographic representation (3) miniature scale model not allowing manipulation of parts and materials and (4) miniature scale model allowing manipulation of parts and materials. Teams of five men and individuals as individuals were required to solve the problem at each of these four levels of presentation and write out their solutions.

The evidence indicated that the written solutions at the four laboratory levels of remoteness from reality for teams and for individuals were equivalent, i.e., not statistically different from each other. The written solutions by teams, however, were markedly superior to those by individuals at each of the four laboratory levels of presentation. The apparent equivalence in goodness of solutions among the laboratory levels of presentation was attributed in part, at least, to the fact that teams and individuals at each laboratory level were given the opportunity to equalize the amounts of information about all aspects of the problem, either by perception directly or by questioning indirectly. The superiority

---

[1] Studies in Reality. No. 2. Conducted under contract Number AF 18 (600)-341 of the Human Resources Research Institute, Maxwell Air Force Base, by the Institute of Psychological Research, Teachers College, Columbia University, Principal Investigator: Irving Lorge.

of team solutions over those of individuals, to a degree, may be attributable to the fact that teams of five members asked more questions than did single individuals, thus, increasing the likelihood of attaining better solutions because of more information.

The laboratory settings in which the Mined Road Problem had been presented were quite remote from actual reality. The question that must be asked is whether differences would be obtained between settings more closely approximating reality and whether variations in conditions approximating reality would be significantly different from each other. If the Mined Road Problem were developed in actual reality in the field and individuals and teams were required to solve it in such a setting, it is likely that solutions superior to those in the laboratory would come because of the more direct knowledge by direct perception. If the field setting were varied so that the amount of direct information by perception could be changed, then, it is quite likely that such variations would introduce differences in the solution. For instance, one variation may be to allow the subjects just to see the actual field situation without touching any of the equipment around it. In such circumstances, less direct information may be obtainable. Or, again, the subjects may be allowed to manipulate any or all aspects of the materials but not allowed to carry a solution to completion. It is possible that they would get more information by the direct manipulation but less than if they had completed the solution. Therefore, it is hypothesized that there will be a genuine difference in the quality of solutions between those made in laboratory settings and those made in field settings and, moreover, among the several possible levels of field presentations there would be differences attributable to differences in the amounts of information obtainable directly by perception.

The Mined Road Problem was constructed in reality in open country. Three variations were possible with a real field setting:

1) The Actual Real Without Manipulation, where men were not permitted to manipulate the beams, tire, ropes, etc.

2) The Actual Real With Manipulation, where the men were permitted to manipulate all equipment but not to carry a solution to completion, and

3) The Actual Real Solve, where the men were actually required to carry the solution to completion, i.e., to actually cross the road.

The description of the problem was identical with that used in the earlier study, except for such modifications as when necessary to meet the conditions of the field levels of presentation. For the actual real without manipulation, the description was as follows:

> You men are on your way back from a sabotaging mission in enemy-occupied territory. You have just blown up a bridge about a mile from here. According to a prearranged plan, you will meet a guerrilla truck about a mile away from here which leaves you only a very short time to get across this road. You have discovered that this road has been mined with a new type of enemy mine which is supersensitive and will blow up if anything touches it. The explosion at the bridge has aroused the enemy but as yet they do not know in what direction you are going. Your problem is to get the entire group of five men across the road and to leave as little trace as possible of your escape route. You must cross the road at this spot, which is marked off by the white tape.
>
> [RO] In determining your plan of action, you may observe but not manipulate any equipment in the working area, which is marked off by the white tape.
>
> After you have decided how you are going to cross the road, one man is to write your plan of action on the paper provided.
>
> You may ask any questions you wish and we will try to answer them as best we can. Remember, it is important that you complete this mission as quickly and with as little loss of men and material as possible.
>
> According to your schedule you must meet the guerrilla truck in one hour. Remember, also, that your performance on this mission will be compared with that of other members of Air Force Reserve Officers Training Corps personnel.

For the actual real with manipulation, the following paragraph was substituted for paragraph RO:

> In determining your plan of action, you may manipulate any equipment found here in the working area, which is marked off by the white tape. However, at no time may any man or equipment touch or extend over the road.

For the actual real solve, paragraph RO was omitted.

The subjects were cadets of the Air Force Reserve Officers Training Corps at Manhattan College, a corps of men fairly homogeneous in background and intelligence and highly motivated to work with tactical problems. Ten teams of five men and ten individuals as individuals, randomly selected and assigned, were set the task of solving the problem at each of the three field levels. If the problem was given for team solution, the five team members were brought as a group to the field and each given a typed problem to read to themselves while the examiner read it aloud. If the problem was presented for individual solution, only that person was brought to the field, given the typed description to read to himself while the examiner read it aloud. Teams and individuals were permitted and encouraged to ask questions.

For the actual real without manipulation and for the actual real with manipulation, teams and individuals proceeded with their interrogations and deliberations until a solution to the problem was ready to be written. Then either the individual or a member of the team wrote the solution. For the actual real solve, the solution was written by the individual or by a team member after the road crossing had been accomplished. Thus, sixty written solutions were obtained, ten by teams and ten by individuals for each of the three field presentations.

The sixty written solutions were evaluated by Quality Point Score (QPS), i.e., the aggregate of separate weights for each element in the solution. The weights for each element were estimated, giving consideration to factors, such as safety, efficiency, workability, and quality of thinking. The QPS centered around four major elements: (1) Bridging the road, i.e., the use of beams or a combination of beams, swinging, broadjumping, throwing men or materials across, making an overhead bridge, etc. (2)Removing the bridging from the road, i.e., any procedure for removing the basic bridging or secondary bridging. (3) Removing clues indicative of escape or escape route, i.e., materials hidden or carried away. And (4) time, i.e., elapsed time from presentation of the problem to the formal written solution. The written solutions for the three field levels of presentation contained some new elements which did not appear in the solutions at the laboratory levels more remote from reality in the earlier study. Each such new element was classified properly and assigned a reasonable numerical weight.

For teams and for individuals, at each of the three field levels of presentation, the means and standard deviations of the QPS are given in Table 1. The analysis of variance among levels of presentation and between teams and individuals is presented in Table 2.

No evidence of significant differences among the solutions at the three field levels of presentation was obtained either for teams or for individuals. At every field level of presentation, however, the solutions by teams were statistically superior to those by individuals. The apparent equivalence among field levels of presentation, as well as the significant differences between teams and individuals, are like those reported in the earlier study.

TABLE 1.—QUALITY POINT SCORE FOR MINED ROAD SOLUTIONS: MEANS AND STANDARD DEVIATIONS, BY TEAMS AND BY INDIVIDUALS AT EACH OF THE THREE FIELD LEVELS OF REMOTENESS FROM REALITY: TEN TEAMS AND TEN INDIVIDUALS AT EACH LEVEL

Level of Remoteness from Reality	Teams		Individuals	
	Mean	SD	Mean	SD
Actual Real without Manipulation	44.8	29.0	11.9	14.3
Actual Real with Manipulation	54.2	29.1	15.5	18.7
Actual Real Solve	57.9	16.4	39.6	25.2

TABLE 2.—QUALITY POINT SCORE FOR MINED ROAD SOLUTIONS: ANALYSIS OF VARIANCE BETWEEN TEAMS AND INDIVIDUALS AND AMONG THE THREE FIELD LEVELS OF REMOTENESS FROM REALITY

Source	Sum of Squares	Degree of Freedom	Mean Square	F Ratio
Class (Teams and Individuals)	13470.0	1	13470.0	23.1*
Treatment (Level of Reality)	4344.1	2	2172.1	3.7
Class X Treatment	1105.0	2	552.5	
Error	31499.9	54	583.3	
Total	50419.0	59		

* Significant at the .01 level.

Although no significant differences were found among the solutions at the four laboratory levels of presentation in the earlier study and at the three field levels of presentation in the present study, the question may be raised whether there are differences between levels of presentation in the laboratory and those in the field. To answer this question, the analysis of variance was made for QPS among the seven levels of presentation. The analysis showed that there were no significant differences in QPS among the seven levels of presentation, which means no differences between laboratory and field.

A qualitative analysis of the elements in the solutions showed some differences between solutions by teams in the laboratory and field. Of the teams solving the problem in the field, twenty per cent used a diving board method to get men from the starting, to the other side of the road; but, no teams employed this method in the laboratory setting. This diving board involved placing the beam on a fulcrum, pushing the beam partway over the road so that one or more men could walk to the end of the beam and jump to the far side, while the remaining men hold down the base end of the beam. Twenty per cent of teams in the field used swinging to get men across the road. Although the same proportion of teams used swinging in the laboratory, there are differences in how the swinging was done. In the field, all but one team used swinging in conjunction with bridging with a beam. In the laboratory, all but one team used swinging as the only way of crossing the road, or to set up an overhead cable. Only seventeen per cent of field teams, compared with fifty per cent in the laboratory, used an excessive amount of time to solve the problem, thus failing in that they would not meet the guerrilla truck. For individuals, there was no difference between field and laboratory settings.

Differences between teams in the field and those in the laboratory may be due to differences in perception of the problem situation. Despite the differences described above, the elements used in the written solutions were basically the same for the laboratory and field.

It was hypothesized that levels more closely approximating actual reality in a field setting would be differentiated from presentations in laboratory settings which are quite remote from reality. The rationale for the hypothesis is that more information can be obtained directly through perception in the actual field settings than indirectly by questioning in laboratory settings. The facts indicate that there are no significant differences in solutions between laboratory and field levels of remoteness. Moreover, the hypothesis that there would be differences among the three field levels of presentation because of the differences in the amounts of information attainable by direct perception was not supported: none of the differences found among the three field levels of remoteness were significantly different from each other. Indeed, there were no significant differences among all seven levels of presentation.

The finding maintained at each level of presentation, both field and laboratory, was that groups are superior to individuals in the quality of solutions written. Although the evidence in this study supports the fact that groups are superior to individuals, further investigation is necessary to ascertain whether the generalization about group superiority is attributable to interaction among group members, or, as has been suggested by Lorge and Solomon, attributable to the greater probability of getting at least one solver in any group. Research on this phase is now in progress.

REFERENCES

1. I. Lorge, J. Tuckman, L. Aikman, J. Spiegel, and G. Moss. "Solutions by teams and by individuals to a field problem at different levels of reality" *Journal of Educational Psychology*, Jan., 1955, pp. 17–24.

2. The OSS Assessment Staff, *Assessment of Men*. New York: Rinehart and Co., Inc., 1948.

PUPILS' VALUES AND THE VALIDITY OF THE MINNESOTA TEACHER ATTITUDE INVENTORY

G. M. DELLA PIANA AND N. L. GAGE[1]

Bureau of Educational Research, College of Education

University of Illinois

Does the degree to which a characteristic of a teacher affects his acceptance by a pupil depend on what the pupil wants to get out of the relationship? That is, do the values of the pupil determine what characteristics of the teacher will influence the pupil's perception of him? If so, and if the characteristics of the teacher are considered to be determiners of the pupil's evaluation of him, then the values of the pupil may be viewed as 'second-order determiners' —i.e., determiners of what characteristics will be determiners.

'Second-order determiners' are akin to the kinds of research variables that have been espoused by Edwards and Cronbach (*8*) and by the Committee on Criteria of Teacher Effectiveness of the American Educational Research Association (*1*, pp. 253–254). The former writers recommended that research on methods of psychotherapy take into account characteristics of the clients, or 'organismic variables'. The AERA Committee referred to such characteristics of pupils, which might change how a given teacher affected the pupils, as 'intervening variables'.

Similarly, social psychologists have conceptualized leadership as a relation between leader and follower. How this approach has developed is described by Newcomb (*15*, pp. 650–663), and its implementation in research on leadership is illustrated by Sanford (*16*, pp. 328–340). The latter's study shows how the follower's attitudes, preferences, and needs tell us about his 'readiness for leadership'.

---

[1] A more complete presentation of the data and instruments of this study is available in G. M. Della Piana, "Cognitive-Affective Values of Pupils and Teacher-Pupil Relationships," unpublished Master's thesis, 1953, on file in the University of Illinois Library. The thesis was done at the suggestion and under the direction of the second author. A grant from the Bureau of Educational Research, College of Education, University of Illinois, supported the research. We are grateful to Professors Lee J. Cronbach and David R. Krathwohl for a critical reading of the manuscript.

In this study, we consider the teacher as a leader and we deal with one aspect of the teacher's effectiveness: how his pupils rate his competence. This study emphasizes the values of the pupil as factors in teacher effectiveness. It was partly suggested by an investigation (*9*) in which the correlation between the Minnesota Teacher Attitude Inventory (MTAI) scores of twenty teachers and ratings of the teachers by their students was −.18, i.e., negative, but not significantly different from zero. How could this result be assimilated to other findings (*2*, *3*, *11*, *12*, *13*) of significant positive relationships? The two hundred students in the Gage-Suci study were the entire enrollment of a university high school. Impressionistic evidence suggested that these students had relatively strong 'cognitive' values with respect to teachers. That is, they wanted teachers with high 'cognitive merit' who could help them satisfy their 'cognitive', or knowing and understanding, purposes as distinguished from their 'affective' needs for security and acceptance. Such 'cognitive' values could render the affective components of teacher competence less important to the students.

METHOD

Test Materials.—Teacher attitude was measured with the MTAI. This instrument was constructed to measure attitudes related to ability to get along with pupils (*3*). In our view, these attitudes bear particularly on the teacher's 'affective merit'—or effectiveness in satisfying the emotional needs of pupils for security and acceptance. It was scored with the published 'rights' minus 'wrongs' keys.

General pupil ratings of teachers were obtained by use of the fifty-item "My Teacher" inventory constructed by Leeds (*11*, *12*) to appraise pupils' attitudes toward their teachers. This inventory was used to provide an exact replication of previous validations of the MTAI. The questions in this inventory deal with the teacher's disposition, his interest in children's activities, his status in the pupils' esteem, his sense of humor, and similar matters pertinent to teacher-pupil relations in the pupils' frame of reference. A rating for each teacher was obtained by scoring each pupil's Yes-No-? responses (number of favorable minus the number of unfavorable ratings of the teacher) and then computing the mean over pupils.

Pupils' ratings of logically defined characteristics of teachers were obtained with an extension of Leeds' "My Teacher" inventory. To test our hypotheses we needed three additional inventories, for logically more specific kinds of teacher merit. A mean rating for each teacher was obtained for each of these three measures by averaging pupil scores consisting of the number of favorable responses. A brief description of these measures follows:

A) *General Merit*, or the extent to which a teacher is liked or disliked, without specification of any particular kind of reason for the like or dislike. Two of the eleven items of this inventory follow:

57. Would you like to move to a different teacher if you could?........................Yes No ?
63. Are most of the pupils glad to be in this teacher's class?......................Yes No ?

B) *Cognitive Merit*, or the extent to which a teacher is seen as effective in helping the pupil achieve the traditional cognitive, intellectual, subject matter objectives of school learning. Two of the twelve items in this inventory follow:

6. Does this teacher explain school work so that you can understand it?.................Yes No ?
56. Do you learn a lot of things from this teacher?. Yes No ?

C) *Affective Merit*, or the extent to which a teacher is seen as effective in helping the pupil satisfy his social and emotional needs, especially through providing a warm and supportive personal relationship. Two of the eleven items in this inventory follow:

3. Does this teacher, whose name is written above, praise you for doing good work?.........Yes No ?
4. Does this teacher scold the pupils a lot?.....Yes No ?

These three logically defined inventories contained a total of thirty-four items, obtained as follows: Fifteen of the thirty-four items were written by the authors to fit the definitions of the categories; the other nineteen items were selected from among those already available in Leeds' "My Teacher" inventory because they seemed to fit our definitions of the categories. Eleven judges (educational psychologists) were given the thirty-four items to classify according to the three definitions. At least eight of the eleven judges agreed on the classification of all thirty-four items; on seventeen of the items, the judges were unanimous.

Pupils' cognitive-affective values with respect to teachers were measured with a forced-choice inventory ("Choosing Things About Teachers") which was designed to reveal how much pupils value the two kinds of teaching merit (cognitive and affective) involved in our hypotheses. This inventory included thirty-six items, each containing two phrases—one describing cognitive merit and one describing affective merit. Pupils were asked to check the phrase in each pair which described the quality they wanted more in a teacher. A sample item follows:

Which do you want more? A teacher who

1. doesn't make us afraid or,
2. explains so we can understand.

Pupil scores consisted of the number of 'cognitive choices' made. A mean cognitive-affective value score was obtained for each class. This mean score represents the kind of teacher a given class tends to value and is not a rating of the actual teacher. Due to the forced-choice nature of the test, a high class mean indicates high cognitive value orientation, while a low class mean indicates high affective value orientation.

Subjects.—Ninety-seven of the ninety-eight fourth-, fifth-, and sixth-grade classes in a mid-western city of 66,000 population participated in the study. This group, consisting of ninety-seven teachers and their 2,704 pupils, is similar to the groups used in validation studies by the authors of the MTAI. Class size ranged from six to thirty-nine pupils, with a median of twenty-six. Pupils and teachers were assured that individuals and schools would not be identified.

Administration.—All devices were administered by graduate students from outside the school system. A detailed manual of directions was followed to maximize uniformity of procedure. The teachers took the MTAI in another room while their pupils were filling out the various inventories. All pupils finished in less than fifty minutes. The tests were given in late spring, after pupils had been with their teacher for more than six months.

RESULTS

Characteristics of Tests.—(Reliability) The reliabilities of the MTAI, Leeds, Affective merit, and General merit inventories were all above .90. The Cognitive merit reliability estimate was .83. Guttman's L_4 coefficient (*5*) was used for the MTAI estimate, and Horst's coefficient (*10*) was used for the other estimates. These measures appear to differentiate among teachers and classrooms with sufficient accuracy for testing our hypotheses.

Two estimates of the reliability of the Cognitive-Affective value scores were obtained: A Kuder-Richardson Formula 21 estimate of .89 and a Horst estimate of .74. While this K-R 21 coefficient indicates that our measures of the Cognitive-Affective Values of individual pupils are internally consistent, it is irrelevant when we are using mean scores of a class to test our hypotheses. Because the Horst coefficient for Cognitive-Affective Value class means was only .74, we used only the classes at the extremes of the distribution of these means for testing hypotheses involving Cognitive-Affective Value scores. Hence, in one method of testing Hypothesis II, below, we used only the twenty classes with the highest mean Cognitive Value scores and the twenty classes with the highest mean Affective Value scores. For the other method described below, the K-R 21 coefficient of .89 appears adequate.

(Intercorrelations) As noted below, one of our hypotheses depends upon distinctions between logically defined types of teacher merit as judged by pupils. Unless these measures are substantially independent, we cannot use them to test hypotheses concerning the differential relevance of the MTAI to teacher effectiveness. The intercorrelations shown in Table 1 indicate how our logical distinctions held up empirically. The r between the logically and operationally independent Affective and Cognitive Merit scores is .68. Taking into account the reliabilities and intercorrelation of these mean pupil scores, we find, using Bennett's monograph (*4*, p. 152), that only about twenty-one per cent of the differences in these test scores may be considered in excess of chance expectation. When corrected for attenuation, their intercorrelation becomes .79. Tests of hypotheses involving use of these scores as differential measures are therefore unlikely to yield significant results; we are especially likely to find the null hypothesis tenable when it is in fact false. This overlap does not, however, preclude using these scores as measures of pupils' general approval of teachers.

The Leeds inventory asked pupils to describe their own teacher; the values inventory asked them to check which characteristic of a teacher they would prefer if they had to choose between a 'cog-

TABLE 1.—CORRELATIONS AMONG MTAI AND MEAN PUPIL SCORES OVER ALL TEACHERS (N = 97)

Source	2	3	4	5	6
1. MTAI	.256*	.210*	.262**	.248*	.04
2. Leeds' "My Teacher" Inventory		†	†	†	.05
3. Affective merit			.68**	.84**	.10
4. Cognitive merit				.70**	.08
5. General merit					.03
6. Cognitive value					

* Significant at .05 level

** Significant at .01 level.

† Since scores 3, 4, 5 are based in part on items also contained in score 2, their correlations with the latter are omitted.

nitive' and an 'affective' virtue. Table 1 shows that the correlations of the means of the cognitive value scores with the means of each of the ratings of teachers are very low; our inventory of pupils' values measured something different from pupils' attitudes toward their own teacher.

TESTS OF HYPOTHESES

Hypothesis I. Teachers' attitudes as measured by the MTAI correlate positively with mean pupil ratings of teachers on the Leeds inventory. Our obtained coefficient of .26, significant at the .05 level, corroborates previous studies with similar teacher-pupil populations (*12*, *13*). That is, the teachers' MTAI scores have a significant positive correlation with their concurrent ratings by their pupils on the Leeds inventory.

As shown in Table 1, the correlations between teachers' MTAI scores and their Affective merit, Cognitive merit and General merit ratings were all positive and significant at or beyond the .05 level. Since these scores were highly correlated with Leeds scores they simply offer additional evidence for a significant positive correlation between teachers' MTAI scores and pupils' general approval of teachers. From our earlier discussion, however, our prediction is that this relationship will change when pupil value orientation is taken into account. Our next hypothesis bears on this point.

Hypothesis II. Teachers' MTAI scores correlate with pupil liking for teachers more highly where pupils value more highly the social-

TABLE 2.—CORRELATIONS BETWEEN MTAI AND PUPIL RATINGS FOR TEACHERS GROUPED ON THE BASIS OF THEIR PUPILS' MEAN COGNITIVE-AFFECTIVE VALUE SCORE

Class Cognitive Value Standing	N	Pupil Ratings Correlated with MTAI			
		General merit	Cognitive merit	Affective merit	Leeds rating
Highest	20	.24	.15	.05	.05
Middle	57	.16	.18	.20	.24
Lowest	20	.59**	.58**	.50*	.57**

* Significant at .05 level.
** Significant at .01 level.

emotional need mediating behavior of teachers. Specifically, for each affective merit-type measure (general merit rating, affective merit rating, and Leeds "My Teacher" rating) the correlation between MTAI score and affective merit-type measures for the most affectively oriented pupils will be greater than that for the most cognitively oriented pupils. In brief, we hypothesize that:

$$r_{\substack{\text{(MTAI vs. affectively valued pupils'}\\ \text{ratings)}}} > r_{\substack{\text{(MTAI vs. cognitively valued pupils'}\\ \text{ratings)}}}$$

Two methods of analysis were used to test this hypothesis.

IIa. Here we divided pupils into high and low cognitive value groups on the basis of the mean score of their class. Table 2 summarizes results of this analysis. The differences between r's are all in the hypothesized direction. Although not hypothesized, the correlation between MTAI scores and cognitive merit ratings also changes in this direction. In view of the high correlation of cognitive merit ratings with affective merit, general merit and Leeds ratings, this result is to be expected. The difference between high cognitive and high affective clăsses in the correlations of MTAI scores with Leeds ratings is significant at the .05 level. Here the r for high cognitively oriented classes is .05, while that for low cognitively oriented classes is .57. The test of significance used here was between two z-transformations of r, for $N_1 = N_2 = 20$; since the direction of the difference was hypothesized in advance, a one-tail test of significance was used.

This significant difference might be due to differences in the heterogeneity (or 'range of talent') of these scores in the two groups.

TABLE 3.—RELIABILITY OF MTAI AND LEEDS SCORES IN LOW AND HIGH COGNITIVE VALUE GROUPS

Cognitive Value Group	N	Mean*	Range	SD	Reliability**
			MTAI score		
Highest	20	72.00	12 to 115	29.8	.90
Lowest	20	61.75	−36 to 106	37.1	.92
			Leeds score		
Highest	20	19.28	−9.16 to 36.04	9.79	.93
Lowest	20	16.02	−5.31 to 30.04	9.87	.90

* On both scores, the means of the two groups are significantly different at the .001 level of confidence.

** For the MTAI this is Guttman's L_4 coefficient (*5*); for the Leeds score, this is Horst's coefficient (*10*).

Table 3 presents data bearing on this possibility. The differences in the variances and reliability coefficients were not significant. In fact, the only significant differences noted between the high and low cognitive value groups are differences between the means for each group of teachers on the MTAI and Leeds devices. These differences have no bearing on the correlations under consideration. So differences in group heterogeneity with respect to the variables under consideration do not account for the difference between the correlations of the MTAI with the Leeds scores in the high and low cognitive value groups. Indeed, when the correlations were corrected for attenuation, they differed even more than previously; the corrected correlation between MTAI and Leeds score for the high cognitive group was .05; for the low cognitive group, the corrected r was .63.

IIb. A second method of testing Hypothesis II is the following: the correlation between teachers' MTAI score and affective merit-type rating based on the most affectively oriented pupils in each class will be greater than that based on the most cognitively oriented pupils in each class. Here, Hypothesis II takes the form:

$$r_{\substack{\text{(MTAI vs. rating by most affectively}\\ \text{oriented pupils in each class)}}} > r_{\substack{\text{(MTAI vs. rating by most cognitively}\\ \text{oriented pupils in each class)}}}$$

In this method of testing Hypothesis II, we divide the pupils in

TABLE 4.—CORRELATIONS BETWEEN MTAI AND PUPIL RATINGS FOR PUPILS GROUPED ON THE BASIS OF THEIR COGNITIVE-AFFECTIVE VALUE SCORE

Group of Pupils in Each Class According to Cognitive Value Score	No. of Groups and of Teachers	Leeds ratings			MTAI			r (Leeds vs. MTAI)	r Leeds vs. MTAI, corrected for attenuation)
		Mean of means	SD of means	Reliability (Horst)	Mean	SD	Reliability (L_4)		
Above median of their class	97	22.22	9.85	.83	56.80	35.63	.90	.21	.24
Below median of their class	97	19.36	11.97	.85	56.80	35.63	.90	.25	.29
All pupils with 21 or above	64	20.66	10.91	.80	52.03	33.86	.90	.29	.35
All pupils with 15 or below	64	17.27	13.28	.79	52.03	33.86	.90	.35	.42
5 pupils with highest scores, of those 21 or above	64	23.52	10.56	.52	52.03	33.86	.90	.21	.31
5 pupils with lowest scores, of those 15 or below	64	16.62	15.13	.74	52.03	33.86	.90	.27	.33

each class on the basis of their cognitive-affective values. We then compute two mean Leeds ratings for each teacher: one for the most affectively oriented pupils and another for the most cognitively oriented pupils. The teachers' MTAI scores are then correlated against each of the two sets of mean ratings.

Table 4 shows the results obtained by three procedures for dividing the pupils in each class into high and low cognitive value groups. The first procedure was to divide the pupils in each class into groups that were above and below the median cognitive value score of their own class. For the pupils above the median, the correlation of their Leeds ratings and their teachers' MTAI scores was .21; for pupils below the median this r was .25. Corrected for attenuation due to the imperfect reliability of the mean Leeds ratings and MTAI scores, they become .24 and .29, respectively.

The second procedure for dividing the pupils in each class was to choose all pupils who had a cognitive value score of 21 or above for the high cognitive value group, and all those with scores of 15 or below for the low cognitive value group. Since not all classes had at least five pupils with scores above or below these cutting points, we were able to use only sixty-four classes in this analysis of the data. Here the r for the high cognitive value group was .29, while that for the low cognitive value group was .35. Corrected for attenuation in both MTAI and mean Leeds ratings, the coefficients became .35 and .42, respectively.

The final procedure for dividing the pupils was to use only the five pupils with the highest scores, of those pupils who had cognitive value scores of 21 or above, and only the five pupils with the lowest scores, of those that had scores of 15 or below. This was done to make the numbers of pupils the same in the high and low cognitive value groups. This would also make the r's between Leeds ratings and MTAI scores more comparable. The correlations obtained with these kinds of groups were .21, for the high cognitive group and .27 for the low cognitive group. Corrected for attenuation these r's become .31 and .33, respectively.

Obviously, the differences between r's by each of the three procedures within this second method of analysis are not nearly so large or significant as that obtained by the first method of analysis, where the twenty highest and lowest cognitive value classrooms were used intact. The differences are all in the hypothesized direction, by this second method of analysis. They are probably not statistically significant although a suitable method of testing the significance of differences between such r's is not available (*14*, pp. 217–218).

How might we explain the differences in results between these two methods of analysis? However speculative and a posteriori, the following two arguments may be considered: First, it is possible that the intact classroom must be the unit of analysis. In breaking up classes into high and low cognitive value sets of pupils, we may be losing sight of the possibility that, when pupils rate their teachers, they are influenced not by their own individual values but by their perception of the entire class's (mean) values. In other words, the disagreement in results between our two methods of analysis suggests that the entire class's mean of pupils' values may have provided a frame of reference for the entire class and influenced their Leeds ratings accordingly. Breaking up the class into high and low cognitive value groups would not destroy the influence of this class-wide frame of reference.

Also, the less definitive results of the second method of analysis may be accounted for by the fact that the Leeds ratings of each teacher are in this instance based on fewer pupils. This makes the mean ratings less reliable; hence the two correlations of these mean ratings with the MTAI both regress toward zero and become more equal. The correction for attenuation applied to the two more equal r's, since it raises both r's about equally, cannot then do justice to any hypothesis involving differences between r's.

At any rate, the two methods of analysis agree in the direction of the differences between r's yielded by them. In this sense, Hypothesis II is supported, at the .05 level by the first method of analysis, but without statistical significance by the second.

Hypothesis III. Teachers' MTAI scores correlate positively with the cognitive value orientation of their classes.—This hypothesis is based on two assumptions: (a) A teacher with a higher MTAI score is likely to be more successful in satisfying his pupils' social-emotional needs; his pupils will tend less to need or want this kind of need-satisfying behavior from a teacher. (b) With social-emotional needs satisfied, other needs become dominant; in the classroom, the need to learn 'subject matter' is likely to become the dominant one for most pupils.

The negligible correlation between MTAI and cognitive value (r = .04) does not, however, support this hypothesis.

SUMMARY

Current theory suggests that leadership consists of an interaction between leaders' characteristics and followers' values and needs. Within this framework, we hypothesized that pupils' liking of teachers is a function of the interaction between pupils' values and teachers' attitudes. This means that the significant positive relationship (corroborated in this study) between teachers' scores on the Minnesota Teacher Attitude Inventory and ratings of the teachers by their pupils, should change predictably as pupils' value orientations are varied. A forced-choice values inventory was constructed to measure how much pupils want teachers with cognitive merit (effectiveness in helping pupils achieve intellectual objectives of school learning) as against affective merit (effectiveness in helping pupils satisfy their social-emotional needs). For the pupils who have stronger affective values, the teachers' MTAI scores do correlate more highly with how much they are liked by the pupils.

Accordingly, the MTAI will vary in validity for teacher effectiveness according to the values of the pupils interacting with the teacher. Teachers scoring high on the MTAI will probably be better liked by pupils who have strong affective values concerning teachers. If the pupils have strong cognitive values, the teacher's MTAI will make less difference.

The study supports the validity of the interactional point of view in the study of teacher-pupil relationships, or leadership in general. Attitudes and similar characteristics of teachers depend for their significance on the values, needs, and other characteristics of their pupils.

REFERENCES

1) A. S. Barr and Others. "Report of the Committee on the Criteria of Teacher Effectiveness." *Rev. educ. Res.*, 1952, 22, 238–263.

2) R. Callis. "The efficiency of the Minnesota Teacher Attitude Inventory for predicting interpersonal relations in the classroom." *J. appl. Psychol.*, 1953, 37, 82–85.

3) W. W. Cook, C. H. Leeds and R. Callis. *Minnesota Teacher Attitude Inventory: Manual*. New York: Psychological Corporation, 1951.

4) L. J. Cronbach. *Essentials of Psychological Testing*. New York: Harper, 1949.

5) — — —. "Coefficient alpha and the internal structure of tests." *Psychometrika*, 1951, 16, 297–334.

6) W. J. Dixon and J. F. Massey, Jr. *Introduction to Statistical Analysis*. New York: McGraw-Hill, 1951.

7) R. L. Ebel. "Estimation of the reliability of ratings." *Psychometrika*, 1951, 16, 407–424.

8) A. L. Edwards and L. J. Cronbach. "Experimental design for research in psychotherapy." *J. clin. Psychol.*, 1952, 8, 51–59.

9) N. L. Gage and G. Suci. "Social perception and teacher-pupil relationships." *J. educ. Psychol.*, 1951, 42, 144–152.

10) P. Horst. "A generalized expression for the reliability of measures." *Psychometrika*, 1949, 14, 21–31.

11) C. H. Leeds. "The construction and differential value of a scale for determining pupil attitudes." Unpublished doctoral dissertation, University of Minnesota, 1946.

12) ——. "A scale for measuring teacher-pupil attitudes and teacher-pupil rapport." *Psychol. Monogr.*, 1950, 64, No. 6 (Whole No. 312).

13) ——. "A second validity study of the Minnesota Teacher Attitude Inventory." *Elem. sch. J.*, 1952, 52, 398–405.

14) E. F. Lindquist. *Statistical Analysis in Educational Research*. Boston: Houghton Mifflin, 1940.

15) T. M. Newcomb. *Social Psychology*. New York: Dryden Press, 1950.

16) F. H. Sanford. "The follower's rôle in leadership phenomena." In G. E. Swanson, T. M. Newcomb, and E. G. Hartley, (Eds.) *Readings in Social Psychology*. Revised Edition. New York: Holt & Co., 1952.

SCORING DIFFICULTY OF WECHSLER COMPREHENSION RESPONSES

GALEN R. PLUMB

New York State Industrial Institution

West Coxsackie, New York

and

DON C. CHARLES

Iowa State College

One of the problems inherent in individual intelligence testing is evaluating and weighting ambiguous verbal responses to test questions. The Comprehension Subtest of the Wechsler-Bellevue Scale Form I seems to offer particular difficulty to testers. The purpose of this study was to determine the amount of agreement exhibited by a number of judges with varying degrees of training and experience in scoring two-hundred and fifty-four actual responses to the last nine items of the Comprehension Subtest.

The Wechsler test has been widely used in research since its publication in 1939, and has itself been the subject of study. The results of these various studies have been summarized in reports in 1945 (*2*) and in 1951 (*3*).

The Comprehension Subtest of the Wechsler-Bellevue consists of ten questions designed to test 'common sense' (*4*, p. 81). These items are scored by credits of a maximum of 2 points for the best answers, 1 for acceptable responses, and 0 for wrong responses. In administering the test the authors of this study frequently received ambiguous, hard-to-score responses to all but the first item, as for example in item 5.

Question 5: "Shoes" Why are shoes made of leather?
Answers: Don't tear up easy, look good.
Holds better than other material.

This study was designed to determine the extent of agreement and disagreement on such ambiguous responses among examiners of varying experience.

PROCEDURE

Subjects.—A total of two-hundred and fifty-four ambiguous responses were selected from three-hundred and seventy Wechsler-Bellevue scales. These responses were given by a male reformatory population ranging in age from seventeen through twenty-nine years, with a mean age of 21.4 years. The highest school grade completed, as reported by the subjects, was 8.6, and the mean grade placement as determined by the Progressive Achievement test was 7.5. The mean Verbal IQ (W-B I) was 92.4 and the mean full IQ as measured by the California Test of Mental Maturity was 90.5. Most of these men did manual labor before entering the institution. Seventy per cent were classified as unskilled workers, twenty-five per cent were semi-skilled and five per cent were in clerical occupations.

Items.—The original responses were recorded verbatim by the examiners as the subjects were tested. These responses were mimeographed and were given to a number of judges, of varying training and experience, to be evaluated.

Judges.—Two groups of judges were used. The first was an experienced, 'expert' group of twenty-four psychologists. Three had administered one-hundred or more Wechsler-Bellevue scales, two had administered one-hundred and fifty or more, and nineteen had given over two-hundred. Eleven were ABEPP Diplomates in clinical psychology. Six were Associates in the Clinical or Counseling and Guidance Divisions of the A.P.A. Seven were not members of the A.P.A., but were recommended by Diplomates as being qualified authorities in the administration and scoring of this test. The judges worked as professional psychologists in eleven states.

The second group of judges was made up of psychology graduate students with varying degrees of training and experience in administering individual tests.

RESULTS

The scoring of these two sets of judges is summarized in the tables. The experts gave a range of from twelve to thirty-eight per cent of the two-hundred and fifty-four responses a rating of 2; from thirty-five to sixty-seven per cent a rating of 1, and from nineteen to forty-six per cent a rating of 0. The variation among the students is about as great, and the average of each group of judges shows very little difference (Table 1).

TABLE 1.—RESPONSE EVALUATION BY JUDGES

W-B Score	By Experts		By Students	
	Range	Average	Range	Average
	per cent	*per cent*	*per cent*	*per cent*
2	12–38	23	18–31	23
1	35–67	46	26–42	48
0	19–46	31	17–43	29

TABLE 2.—SCORING DISAGREEMENT BY JUDGES

	Expert Less than ⅔ agree	Student Less than ⅔ agree
	per cent	*per cent*
2 Theatre	48	56
3 Company	69	62
4 Taxes	41	41
5 Shoes	40	14
6 Land	32	40
7 Forest	65	56
8 Law	31	38
9 Marriage	48	54
10 Deaf	74	56
Av.	52	49

Table 2 shows the extent of disagreement among the judges in scoring each item. 'Disagreement' was defined here arbitrarily as failure of two-thirds of the judges to agree on scoring an item 2, 1, or 0. Among the expert judges, agreement was not reached for thirty-one per cent (Law) to seventy-four per cent (Deaf) of the responses. The student judges did not reach the agreement criterion for fourteen per cent of the responses (Shoes) to sixty-two per cent (Company). For both groups of judges, lack of two-thirds agreement occurred about half the time on the average.

Further analysis of the data indicates that if the criterion of three-fourths agreement were established, the judges would have failed to agree on sixty-five per cent of the responses; and the percentage of disagreement would increase to eighty-four if seven-eighths agreement were required.

An analysis of the IQ's of subjects in relation to the difficulty of scoring their responses failed to disclose any significant relationship, nor was there any one type of response most puzzling to the judges.

DISCUSSION

It is apparent that the Comprehension Subtest elicits ambiguous and hard-to-score responses. The marked similarity of scoring by experts and graduate students suggests that experience is not of any considerable help in evaluating ambiguous response. Admittedly, it is easier (or at least it might give the tester a feeling of self-confidence) to evaluate the verbal rather than the printed responses of the subject. One of the reasons that response analysis would seem easier in the actual testing process is that additional questioning may elicit clearer answers. However, the amount of questioning permissible may not be clear to either the experienced or inexperienced tester. The manual is not particularly helpful here.

Certain kinds of responses may be mentioned as causing variation in scoring, although these do not account for all the difficulties:

1) A multiple response may be confusing because each part properly should receive a different value. For example, "Keeps government going, keep highways up and such," was one to the "taxes" item. The first part by itself would be a 2-response according to the manual, and the second part would be a 1-response, but together they produce confusion as to the proper score to assign.

2) Words of similar but slightly different meaning may produce confusion in evaluating a response. For example: "Holds better than other material," for the 'shoes' question. Would this response be less confusing if "wears" were substituted for "holds"?

3) A response not similar to those given in the manual may be more confusing than a response similar to those given. For example: "Pick up bad habits," in response to the question "Why should we keep away from bad company?" Does this mean "become like them" as given in the manual?

The above examples might sound easy, and it might be supposed that the correct evaluations would be assigned to them, but the judges in this study divided equally between the 2 and 1 values on these very responses.

It is apparent that these ambiguous responses to the Comprehension Subtest cause difficulty for testers with varying amounts of experience. To aid in alleviating this difficulty, it might be suggested that:

1) The manual for this subtest should contain a wider variety of responses, especially those that might be classified on the 'fringe' of the scoring scale.

2) There should be more specific directions for evaluating a response which requires questioning by the examiner beyond the initial response. Further research on this matter would seem desirable.

3) Difficult responses might be recorded and published together with the evaluations of competent judges. This might be an aid for trainee or experienced tester alike. One such supplementary manual is now available (*1*).

The need for aids to evaluation will continue, since eight of the ten items of this subtest are included in the revision which was being standardized at the time this article was written.

REFERENCES

1) Helen Kitzinger and Eugene Blumberg. "Supplementary Guide for Administering and Scoring the Wechsler-Bellevue Intelligence Scale (Form I)." *Psychol. Monograph*, 1951, 65 (2), No. 319.

2) Arthur I. Rabin. "The use of the Wechsler-Bellevue scales with normal and abnormal persons." *Psychol. Bull.*, 1945, 42.

3) Arthur I. Rabin and Wilson H. Guertin. "Research with the Wechsler-Bellevue test: 1945–1950." *Psychol. Bull.*, 1951, 48, No. 3.

4) David Wechsler. *The Measurement of Adult Intelligence*. Baltimore: Williams and Wilkins, 1944.

AN APPLICATION OF FISHER'S DISCRIMINANT FUNCTION IN THE CLASSIFICATION OF STUDENTS

J. STANLEY AHMANN

Cornell University

In a recent article, Johnson (*3*) briefly described a series of statistical methods of relatively recent development which have decided promise when applied to educational research. One of these techniques mentioned was the discriminant function, introduced by Fisher (*2*), which was designed to determine the linear combination of numerical measurements which would discriminate best the members of one group from those of another group. A direct application of this analysis can be found in the activities of the counselor or personnel officer who desires to weight several characteristics in such a way that maximum discrimination will be obtained between groups of students majoring in two different curricula, or between groups of applicants eligible for two different occupations.

The usage of the discriminant function in educational research has been relatively small, if not nonexistent in many areas. However, a symposium at Harvard University (*4*) reflected considerable interest in the application of the technique, and undoubtedly part of that interest manifested itself in one of the section meetings held at the 1954 Convention of the American Personnel and Guidance Association. At this time, Alman (*1*) reported the usefulness of the discriminant function in the classification of college students. He determined the similarities and dissimilarities in terms of Kuder Interest Inventory scores which existed among major fields of study in the College of Liberal Arts at Boston University.

THE PROBLEM

Another application of Fisher's discriminant function was found in a problem confronting some of the faculty members serving as student advisors at the New York State College of Agriculture at Cornell University. Because of the heterogeneous nature of the curricula included in this college, a rather widely accepted though unofficial classification of these curricula has been used. In the one group, commonly known as the agricultural science group, are listed curricula with such majors as bacteriology, biochemistry, biology, entomology, agronomy, science teaching, botany, and conservation. In a second group, called the general agriculture group, are included major areas of concentration such as floriculture, general farming, poultry, animal husbandry, vegetable farming, agricultural economics, fruit farming, and the like.

Although students are not prevented from transferring from a curriculum in one group to one in the other group, there is a noticeable reluctance on the part of student advisors to recommend such a transfer when a student, who is below average in terms of the standardized test results, proposes to change from a general agriculture curriculum to an agricultural science curriculum. Furthermore, those prospective students who have achieved exceptionally well in high school are sometimes encouraged to enter an agricultural science curriculum rather than a general agriculture curriculum.

The problem was then stated as follows: Can students who enter agricultural science curricula and successfully continue in those curricula be differentiated in terms of available standardized test scores from those who enter general agriculture curricula and successfully continue in those curricula? A sample of one hundred and twenty-eight students was drawn, twenty-nine of whom entered agriculture science curricula and ninety-nine of whom entered general agriculural curricula. All students entered the College of Agriculture in their respective curricula as freshman in the fall term of 1951. In addition, each continued in his chosen curricula up to and including the spring term of 1954, which terminated his junior year. All female students and foreign students were excluded from the sample.

THE FINDINGS

Available for each student as a matter of administrative routine were the total raw scores of the Ohio State University Psychological Test, the total raw scores of the Cornell General Mathematics Test, the speed of comprehension scaled scores of the Coöperative Reading Comprehension Test, total scaled scores of the Coöperative General Achievement Test, Test II: A Test of General Proficiency in the Field of Natural Sciences, and the high school grade-point average in terms of per cent. All of these scores were compiled

TABLE 1.—SUMMARY OF THE DATA

Numerical Variable	Symbol	Means		Point Biserial Correlations	
		Agric. Sci. Students (N = 29)	Gen. Agric. Students (N = 99)		
O.S.U. Psych. Test	X_1	78.3	77.1	0.02	0.3
Cornell Math. Test	X_2	37.6	35.7	0.05	0.5
Coöp. Reading Comp. Test	X_3	60.5	59.3	0.06	0.7
Coöp. G.A.T. Test II: Proficiency in Science	X_4	60.8	57.6	0.10	1.1
High School Grade-Point Average	X_5	87.7	86.1	0.15	1.7

during freshman orientation week and were the basis upon which discrimination was attempted.

The identification of the variables and the mean scores of the two groups are shown in Table 1. Although the difference was usually slight, the agricultural science group consistently surpassed the general agriculture group. The point biserial coefficients of correlation which represented the relationship between each of the numerical variables and the dichotomy were also small. Testing the significance of the difference between each of the correlation values and zero was determined by the formula given by Wert and others (*5*)

$$t = \sqrt{\frac{r^2(N - 2)}{1 - r^2}}$$

In each instance, the t-value was nonsignificant at the five per cent level, although the t-value of the high school grade-point average was greater than that listed for the ten per cent level.

The foregoing conclusion that differences between the two groups could not be demonstrated for each variable individually is not equivalent to saying that it is not possible to discriminate between the two groups in terms of the five variables taken as a group. Because of this and the fact that one group uniformly surpassed the others in terms of the test scores, a discriminant function was computed.

The discriminant function had the general form of:

$$v = a_1 x_1 + a_2 x_2 + a_3 x_3 + a_4 x_4 + a_5 x_5$$

where the x's are the measurements of the different characteristics and the a's are the weights which will produce the maximum distinction between the two groups of students. The values of the a's were found from the simultaneous solution of the following five normal equations:

$$
\begin{aligned}
d_1 &= a_1 \Sigma x_1^2 + a_2 \Sigma x_1x_2 + a_3 \Sigma x_1x_3 + a_4 \Sigma x_1x_4 + a_5 \Sigma x_1x_5 \\
d_2 &= a_1 \Sigma x_1x_2 + a_2 \Sigma x^2_2 + a_3 \Sigma x_2x_3 + a_4 \Sigma x_2x_4 + a_5 \Sigma x_2x_5 \\
d_3 &= a_1 \Sigma x_1x_3 + a_2 \Sigma x_2x_3 + a_3 \Sigma x^2_3 + a_4 \Sigma x_3x_4 + a_5 \Sigma x_3x_5 \\
d_4 &= a_1 \Sigma x_1x_4 + a_2 \Sigma x_2x_4 + a_3 \Sigma x_3x_4 + a_4 \Sigma x^2_4 + a_5 \Sigma x_4x_5 \\
d_5 &= a_1 \Sigma x_1x_5 + a_2 \Sigma x_2x_5 + a_3 \Sigma x_3x_5 + a_4 \Sigma x_4x_5 + a_5 \Sigma x^2_5
\end{aligned}
$$

The d-values were the differences between means of each variable, whereas the sums of squares and sums of crossproduct were the within group deviation values. Substitution of the appropriate values yield, upon solution, the discriminant function:

$$v = -0.00007x_1 + 0.00010x_2 + 0.00012x_3 + 0.00010x_4 + 0.00061x_5$$

To determine whether the two groups of students could be separated on the basis of these five variables, the F-test proposed by Wert and others (*5*) was used:

$$F = \left(\frac{N - m - 1}{m}\right)\left(\frac{k_1k_2}{N}\right) D$$

where N = total number of cases
k = number of cases in a group
m = number of variables
$D = a_1d_1 + a_2d_2 + a_3d_3 + a_4d_4 + a_5d_5$

Solving the equation yielded an F-value of 0.81 which is nonsignificant.

CONCLUSIONS

Since the F-value was nonsignificant, it was concluded that insufficient evidence was found to reject the hypothesis. Thus, successful agricultural science students can not be discriminated from successful general agriculture students in terms of the five variables used. Therefore, should a student desire to transfer from one curriculum group to another, evidence to support or not support the application would have to be found from sources other than the five variables used in this study.

Certainly this information is useful to the student advisors in the College of Agriculture. It is apparent that additional information can be made available to counselors and personnel officials confronted with comparable problems by the application of the discriminant function in the manner illustrated above or in a more detailed treatment such as that used by Alman (*1*).

REFERENCES

1) J. E. Alman. "Cluster Analysis of Major Fields of Study in Boston University College of Liberal Arts." Paper presented at the American Personnel and Guidance Association Convention, Buffalo, New York. April 14, 1954.

2) R. A. Fisher. "The use of multiple measurements in taxonomic problems," *Annals of Eugenics*, 7: 179–188, 1936.

3) P. A. Johnson. "Some promising newer methods," *The Phi Delta Kappan*, 35: 35–40, October 1953.

4) D. V. Tiedeman and Others. "The multiple discriminant function—A symposium," *Harvard Educational Review*, 71–95, Spring 1951.

5) J. E. Wert, C. O. Neidt, and J. S. Ahmann. *Statistical Methods in Educational and Psychological Research*. New York: Appleton-Century-Crofts, Inc., 1954, pp. 364–369.

THE EFFECT OF READING ABILITY ON TWO STANDARDIZED CLASSIFICATION TESTS

CECIL J. MULLINS

Lee College, Baytown, Texas

Most vocational and educational counselors probably wonder, at some time or other, how much timed tests of achievement or aptitude are likely to be influenced by the reading skill of the person taking the tests. Inasmuch as almost all timed tests of aptitude, general mental ability, and achievement require the subject to do some reading which is included in the over-all time for the test, it seems reasonable to expect reading ability to influence performance on the tests.

This paper intends to make a start toward the investigation of the relationship existing between reading ability and timed tests of various attributes. On such a huge project, of course, only a start can be made by a single paper. This paper, then, confines itself to an investigation into the relationship existing between reading ability and two widely used classification tests, the ACE (Q-score only) and the Coöperative Mechanics of Expression, Form A.

GROUP STUDIED AND MATERIALS USED

Group studied.—The group studied in this investigation was a sample of freshman English students at the University of Houston. The correlation between ACE Q-Score and reading speed was obtained on a sample of four hundred students. The correlation between Mechanics of Expression and reading speed was obtained on a sample of five hundred and ninety-nine students.

Materials used.—Reading speed was measured by the speed section of the Coöperative Reading Comprehension Test. The entire test was not used because the investigator wanted to discover to what extent other measures were affected by the Testee's speed of reading, since the other measures studied were timed measures of skill.

The score on the quantitative section of the ACE was used because the investigator believed that if a significant correlation could be obtained on this supposed nonlinguistic section, then it should logically follow that at least as high a correlation could be assumed for the other section and for the total ACE score.

The Mechanics of Expression score was investigated as a sample of measures of achievement in other areas, which, if completely successful, should not be affected significantly by the testee's speed of reading, since the tests are designed to measure proficiency in other areas than reading speed.

TECHNIQUES AND ANALYSIS OF DATA

Pearson product-moment method of correlation.—Pearson product-moment correlations were computed between ACE Q-scores and speed of reading scores and between Mechanics of Expression scores and speed of reading scores, using the scattergram and Garrett's formula 44.[1] This method of correlation yields a correlation between two variables, assuming that the regression lines are linear. This is a very frequent method of correlation used in standardizing tests. The standard errors of the correlations were then computed by Garrett's formula 51,[2] in order to establish the significance of the obtained data.

The results of the correlations obtained by the above method are:

a) Between Mechanics of Expression and speed of reading a correlation (r) of .465 was obtained, with a standard error of .032. This means that the relationship between these two variables is not extremely high, but very significant statistically.

b) Between ACE Q-score and speed of reading, a correlation (r) of .146 was obtained, with a standard error of .049. This correlation is barely significant at the .01 level, using this method in its computation.

The correlation ratio (eta).—The correlation ratio is a method of correlation used for obtaining the degree of correlation between two variables which does not assume a linear relationship. As Garrett points out,[3] this method of correlation could be used for obtaining the degree of relationship in any case in which the Pearson product-moment method of correlation might be used. Actually, the correlation ratio gives us a more exact picture of the degree of relationship. If one has a linear distribution on the two variables, r and eta will coincide exactly. It is so seldom, however, in dealing with human variables, that one obtains an exact linear distribution; it is frequently safer, though somewhat more laborious, to compute a correlation ratio in preference to the Pearson product-moment. Computation of the correlation ratio followed the formula given by Garrett on page 370,[4] corrected by his formula 76.[5] Standard errors were computed using his formula 75.[6]

---

[1] Henry E. Garrett, *Statistics in Psychology and Education*, p. 287.

[2] Ibid., p. 297.

[3] Ibid., p. 367.

Correlation ratios were computed on the same groups mentioned above, with the following results:

a) Between Mechanics of Expression and speed of reading, corrected etas of .739 (representing the change in Mechanics of Expression scores associated with changes in speed of reading scores) and .841 (representing the change in speed of reading scores associated with changes in Mechanics of Expression scores) were obtained, with standard errors of .019 and .012, respectively. This paper is particularly interested in the former of these two etas, since it concerns itself with the effect of speed of reading on Mechanics of Expression. Using eta, then, which is a more exact representation of relationship between these two variables, we have not only a statistically high significance, but also a rather high correlation (much higher than one would expect if the Mechanics of Expression Test measures mechanics of expression and little else).

b) Between ACE Q-scores and speed of reading, corrected etas of .469 (representing the change in ACE Q-score associated with changes in speed of reading scores) and .520 (representing the change in speed of reading scores associated with changes in ACE Q-scores), were computed, with standard errors of .039 and .037, respectively. Again, this paper is particularly interested in the former, and, again using eta, we have not only what is a significant relationship evidenced, but also a fairly high degree of relationship.

SUMMARY AND CONCLUSIONS

In our attempts to measure attributes of achievement and scholastic aptitude, we must be careful not to include in our measurement certain variables which are not controlled, such as the effect of reading speed on timed tests. This study has indicated that a definite and rather high degree of relationship exists between speed of reading and two such measures, quantitative scholastic aptitude and mechanics of expression.

---

[4] Ibid., Formula (73), p. 370.

[5] Ibid., Formula (76), p. 372.

[6] Ibid., Formula (74), p. 371.

A word of caution in regard to interpretation of these data should be mentioned. Although a high degree of relationship obviously exists between these variables, it does not necessarily mean that a causal relationship exists between them. That is to say, although a fairly high degree of relationship exists, one cannot on the strength of the above investigation say definitely that the relationship is brought about by the influence of speed of reading alone. It is entirely possible that some or all of the correlation in evidence can be traced to a third variable, perhaps general mental ability. The correlations obtained are so high it seems hardly likely, but final judgment must be withheld until the results of many other such investigations are available.

THE JOURNAL OF

EDUCATIONAL PSYCHOLOGY

Volume 46 April, 1955 Number 4

TECHNIQUE IN PROBLEM-SOLVING AS A PREDICTOR OF EDUCATIONAL ACHIEVEMENT[1,2]

K. PATRICIA CROSS AND EUGENE L. GAIER

University of Illinois

Successful problem-solving involves both the knowledge of fundamental facts and their effective utilization in arriving at a solution (1; 3). Familiar to teachers and employers is the student or worker who pays little attention to the interrelationships among the facts used. Presumably, he is limited to previously encountered specific situations for which he has memorized the solutions. At the other end, is the individual who has concentrated on an 'understanding' of the general principles underlying the problem. He is not dependent upon rote recall but comprehends and employs the principles involved. No inherent conflict exists, however, between the learning of content and the development of analytical acumen. These processes can be considered not as opposed but as complementary goals of education.

This paper reports the development of a test—the Balance Problems Test (BPT)—designed to assess individual differences in problem-solving. The test affords indices of relative preference for the use of facts versus the use of principles as a technique in problem-solving. The following hypotheses are proposed:

1) Proficiency in solving problems on the test will be positively related to the tendency to use principles rather than facts alone.

---

[1] This research was supported in part by the United States Air Force under Contract No. 33(038)-25726 monitored by Air Force Personnel and Training Research Center. Permission is granted for reproduction, translation, publication, use and disposal in whole and in part by or for the United States Government.

[2] The writers wish to express appreciation to Dr. L. L. McQuitty for his critical judgments and stimulating interest in this study.

2) Both the number of problems solved and the indices of relative dependency upon principles over facts in the BPT will be positively related to measures of educational achievement.

3) Scores on the BPT, reflecting degrees of preference for 'rational' over 'routine' techniques in problem-solving, will compare favorably with tests of 'general aptitude' as predictors of 'problem-solving achievement' (e.g., as indicated by achievement in mathematics).

THE BALANCE PROBLEMS TEST

In the BPT, problems may be solved either through applications of facts[3] for each specific problem or through the use of a general principle applicable to an entire set of problems. The use of principles is assumed to be a more difficult approach than the use of facts, and a slight reward is offered for the use of this method. However, should the subject be either unable or unwilling to put forth the effort required for solution by principles, he is free to use facts. It was thought that a test of this kind would provide a better prediction of achievement in problem-solving through inclusion of an index of preference of method as well as a measure of ability.

Structure of the Test.—(Test design) The Balance Problems Test consists of six sets of problems, each set containing four problems. Each set of problems may be solved by two different methods: (a) use of a single principle, or (b) use of four facts. (The subject may, of course, use some combination of these two methods.) The general principle applies to the entire set of four problems whereas each fact is applicable to only one problem. The solution by facts requires less effort and less time than the solution by a principle. Consequently, it was deemed desirable to word the instructions so that a somewhat higher value was placed on the use of the principle. The subject was to be free, however, to select either method, or a combination of both methods, as he might desire. An incentive for using the principle was provided by informing the subject that one point would be charged for each piece of 'information' used in solving the problem. For each set of four problems, the use of facts carried a penalty of four points whereas only one penalty point would be charged for use of the principle. The penalties were not used directly in scoring the tests, although one of the scores does place a penalty on the use of facts. The primary function of these instructions was to encourage the use of the principle.[4]

---

[3] By 'facts' we mean information which might be utilized in solving a specific problem or a limited category of problems, requiring but simple recall and direct application. By 'principles', we mean information which can be employed in solving a larger class or variety of problems. The use of principles in solving problems requires more complex thought, such as the perception of relationships, the eduction of correlates, as well as a knowledge of the ideas contained in the problem.

(Test format) Each page of the test contains one set of four problems and the information necessary for its solution (hidden from view), as shown in Figure 1.

All sets of problems have the general format shown in Figure 1, but the figures and complexity of the problems change with each set. The sets become increasingly difficult as the subject proceeds through the test.

On the left-hand side of the page, four problems were presented. On the right-hand side of each page of the test is the necessary information for the solution of the problems of a set. The 'Principle' which is applicable to each set is hidden under a tab at the top; beneath are the four facts, all hidden under tabs which must be removed before the information can be read. Once the tab has been removed, it is irreplaceable; the scorers are able to count the number of information tabs seen by the subject. It should be noticed, however, that this does not tell what information was actually employed or how the information was utilized for the actual solution. Many instances were observed in which an effort was first made to use the principle; then, upon failing to solve the problem, facts became the information utilized.

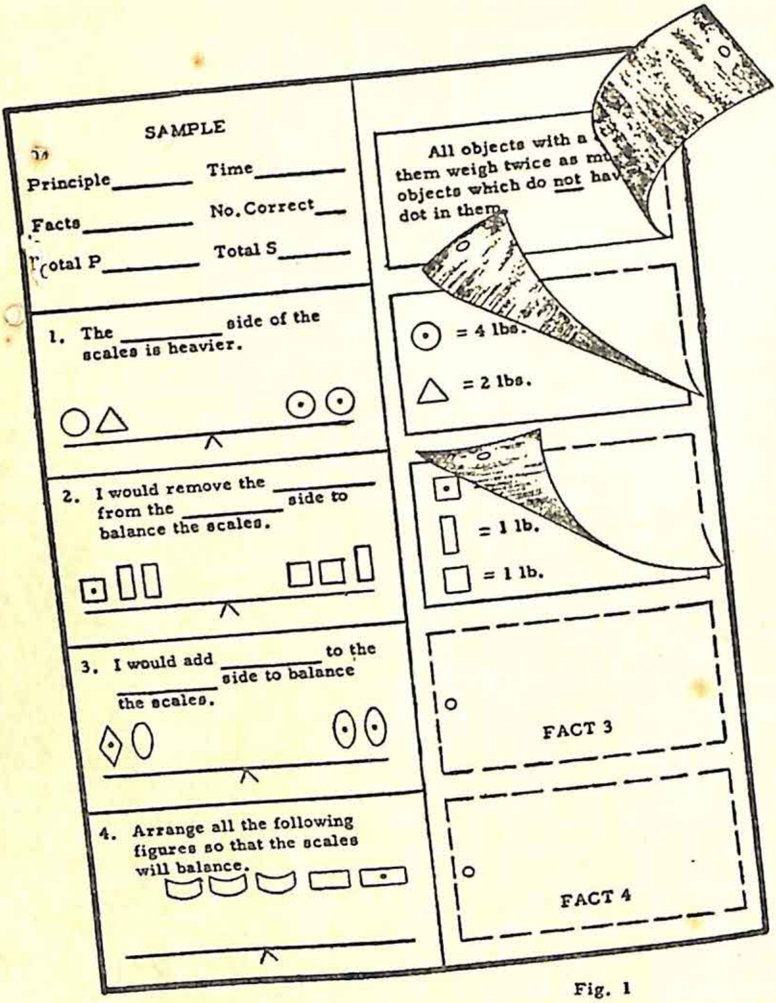
After the tabs had been removed, the principles and facts for the given problem set would appear as shown in Figure 1.

Scoring the Test.—Five scores on the test were derived:

(Variable 1: Number correct.) This score is a count of the number of problems answered correctly irrespective of the kind of in-

---

[4] It is recognized at the outset that these instructions may have seemed to make this test dependent, in part, on individual differences in the effectiveness of the incentive. But, since the incentive was objectively the same for all subjects, individual differences in its effectiveness are an index of the relative strengths of preferences for facts versus principles. Abstract rather than general mechanical principles were employed here so that a minimum of specific knowledge would be required and differences in degrees of training would be minimized.



SAMPLE

Principle______ Time______

Facts______ No. Correct___

Total P______ Total S______

1. The ______ side of the scales is heavier.
2. I would remove the ______ from the ______ side to balance the scales.
3. I would add ______ to the ______ side to balance the scales.
4. Arrange all the following figures so that the scales will balance.

All objects with a … them weigh twice as m… objects which do not hav… dot in them.

Fig. 1

Sample Page of Balance Problems Test

Fig. 1, "Techniques in problem-solving" by Cross and Gaier, pp. 193–206 Apr. Jour. Ed. Psy.

formation used by the subject in his solution. This score was unweighted, the possible range being from zero to a maximum of twenty-four correct.

(Variable 2: Number of facts uncovered.) This score is a frequency count of the number of facts uncovered during the test. No attempt was made here to evaluate the effectiveness of the subject's use of the information uncovered; it is merely the number of 'Fact' tabs torn off, the information being exposed and presumably read. The possible score ranges from zero to a maximum of twenty-four facts uncovered.

(Variable 3: Number of principles uncovered.) This score is a frequency count of the number of principles uncovered during the test. As in Variable 2, no attempt was made to evaluate the subject's effective application of the information uncovered. Rather, this score is merely the number of 'Principle' tabs torn off. The possible score ranges from zero to a maximum of six.

(Variable 4: Time.) Number of minutes used in working the test.

(Variable 5: Weighted test score (S = C + 24 − F).) The weighted score is the number correct + 24 − number of facts uncovered. The constant of twenty-four was inserted to eliminate negative scores in those cases where number of facts uncovered exceeded the number of correct answers. The weighted score was developed as a joint measure of the influence of facts and principles; the fewer facts a subject uncovered, the more he was dependent on principles. The subject who correctly solved all the problems through the use of facts received only half as large a score as the subject who obtained all the correct answers through the use of principles exclusively.

Of the five variables described above, it was thought that the weighted score would be the best predictor of high school mathematics ability. Achievement in mathematics is often presumed to be related to individual differences in the somewhat generalized effective use of principles in solving problems; the weighted score represents an effort to evaluate these differences.

Pre-testing the Instrument.—A preliminary administration of the test was conducted for the purposes of determining approximate testing time, difficulty level, and the comprehensibility of the instructions. For these purposes, the test was administered individually in unspeeded sessions to five high school students and to five airmen who were experienced airplane and engine mechanics.

The test was next given to fourteen experienced airplane and engine mechanics at Chanute Air Force Base, in two group sessions of seven subjects each. This testing resulted in further clarification of instructions and in the revision of some of the problems to a higher level of difficulty. The subjects were carefully observed as they began solving the test problems by the use of principles, as well as when they shifted from principles to the use of facts in their attempt to solve the problems. Upon completion of this pre-test, the test was further revised into the form described above, for use in the present study.

PROCEDURE

Our second hypothesis maintains that relationships exist between BPT scores and student achievement. In order to test this hypothesis, it was necessary to administer the test to a group of subjects for whom appropriate criterion data could be secured. It seemed that such data should include diverse indices of educational achievement for subjects reasonably homogeneous in general social background and educational level. It was possible to secure five 'criterion scores' for the subjects described below; these scores are called 'Variables 6, 7, 8, 9, and 10', with reference to their representation in the correlational analysis.

In addition to the five BPT scores and the five indices of educational achievement, two measures of general aptitude (Language and Non-language Scores) were available for all subjects.

Subjects.—Twenty-five juniors and fourteen seniors (one entire study class) in a high school of eight hundred students served as subjects. The school is located in an Illinois city of approximately 15,000 inhabitants. In addition to BPT scores, the following criterion and prediction data were gathered for this sample:

Criteria.—The selection of criteria of student achievement was governed by considerations indigenous to the test, as well as those unique to the educational background and training of high school subjects. Considerable evidence has been reported concerning the relationship of abstract and relational thinking to mathematical aptitude and training. Since the use of principles in the BPT demanded the ability to generalize a given relationship to a set of problems, it appeared desirable to evaluate the BPT in relation to indices of actual mathematical training and achievement. Thus, a BPT score was to be employed as one of the predictors of the students' previous years' mathematics achievement (i.e., average mathematics grades), and compared with other predictors.

(Variable 6: Mathematics level.) This variable is essentially a keyed score for estimating the highest level of mathematical training reached. The subjects were asked to check on the cover page of the test those mathematics courses they had taken. The list presented contained the following courses: (a) Elementary Algebra; (b) Plane Geometry; (c) Solid Geometry; (d) Advanced Algebra;

(e) Trigonometry; and (f) any College Mathematics. In addition, the subject was asked to list any other mathematics courses studied. It was noticed that the categories and number of mathematics courses taken frequently seemed to represent different levels of mathematics sophistication. For example, it appeared that a differentiation should be made in the mathematics sophistication of the students of general or commercial mathematics and elementary algebra. Likewise, the person taking college mathematics should have reached a level of mathematical functioning not necessarily dependent upon how much high school mathematics he had had. For these reasons, it was concluded that mathematics level, wherein each course was differentially weighted,[5] would be a more sensitive measure of mathematical sophistication than would the actual number of mathematical courses taken. As a check on this, the degree of correlation was determined between mathematical level and number of mathematics courses completed. A product-moment correlation value of 0.98 was obtained. In the subsequent analyses, the mathematics level has been used.

(Variable 7: Average mathematics grade.) The average mathematics grade consisted of averaging the numerical grades for all courses taken. The average was based on anywhere from zero to six grades, depending upon the number of mathematics courses taken. It would have been interesting to have had a measure of the achievement of all subjects at any one level; but, the subjects were never all of the same level in the courses they were taking.

(Variable 8: Yearly Average Grade.) The yearly average grade was obtained by averaging the numerical grades in all the courses taken by the students during the previous year. The number of courses taken by the students ranged from four to six, inclusively. Since the mathematics course grade was included for those who took mathematics, the yearly average grade, in many instances, was not independent of the Average Mathematics Grade (Variable 7).

(Variable 9: Physical Science Achievement.) An index of ability to read and comprehend material in the natural sciences (also, ability to read and interpret graphs and tables). This is a standardized test score obtained from the Illinois Test of Reading Comprehension.

---

[5] Each course was weighted as follows, and a subject's index of mathematical sophistication was a sum of the weights for the courses he took: 0 = no mathematics courses taken; 1 = each course in advanced arithmetic and commercial mathematics; 2 = elementary algebra; 3 = plane geometry; plus one additional point for each of the following courses—advanced algebra, solid geometry, trigonometry; 7 = any college mathematics course.

(Variable 10: Social Science Achievement.) An index of ability to read and comprehend social science material. Like Variable 9, this score was obtained from the Illinois Test of Reading Comprehension.

Other Predictors.—Since the school from which our sample was obtained was one of the participating members of the Illinois Statewide High School Testing Program, standardized indices of aptitude as well as achievement were available.

Two scores of 'general aptitude' were selected for use as predictors, both obtained from the California Short-Form Test of Mental Maturity:

(Variable 11: Language.) An index of ability to manipulate verbal symbols.

(Variable 12: Non-language.) An index of ability to deal with abstract material, other than verbal.

RESULTS

In order to test the hypotheses stated above, product-moment correlations based on the total sample of thirty-nine subjects were computed among all twelve variables utilized in this study. The obtained coefficients are presented in Table 1. It should be noted that for the number of cases used here, product-moment r's must equal or exceed .27 to be significant at the five per cent level, and .37 to be significant at the one per cent level of confidence. Both of these values are based on a single-tailed test, i.e., where the direction of the r is hypothesized, as is the case for all the hypotheses of this study.

The correlation coefficients of Table 1 will be considered in relation to testing each of the hypotheses outlined above; and, in relation to the possibility of yielding other hypotheses for subsequent investigations.

First hypothesis.—According to this hypothesis, proficiency in solving problems on the test is positively related to the tendency to use principles. The measure of proficiency here used is the number of problems of the BPT correctly solved, Variable 1 of the table. The measures of the tendency to use principles are Variable

Table 1.—Product-Moment Correlation Coefficients for Scores on the Balance Problems Test and School Achievement Data*

Variables	1	2	3	4	5	6	7	8	9	10	11	12
1. No. correct												
2. No. facts uncovered	−29											
3. No. principles uncovered	29	−76										
4. Time-minutes	17	−02	20									
5. C + 24 − F	63	−93	73	06								
6. Math. level	54	−61	45	−03	71							
7. Aver. math. grade	28	−46	34	−03	49	33						
8. Yearly aver. grade	28	−44	29	−05	47	42	83					
9. Phys. Sci. Ach.	61	−37	24	−03	54	57	52	45				
10. Soc. Sci. Ach.	32	−60	36	02	61	34	52	55	66			
11. Language score	42	−45	35	−18	53	60	55	63	71	66		
12. Non-lang. score	55	−43	29	−06	57	44	44	34	60	38	52	

* Decimal points have been omitted.

2, number of facts uncovered (an inverted index), and Variable 3, the number of principles uncovered. Each of these variables has a correlation of .29 with number of problems correctly solved. The correlation with the number of facts is negative and the one with number of principles is positive. These two *r*'s are only slightly more significant than the five per cent level, using a one-tailed distribution. These findings barely substantiate the hypothesis that proficiency on the test is positively correlated with the tendency to use principles. They indicate that subjects tend to be more proficient in problem-solving to the extent that they prefer to use principles rather than facts.

Second hypothesis.—According to the second hypothesis, both the number of problems solved and the indices of relative dependency upon principles over facts in the BPT are positively related to measures of educational achievement.

In Table 1, the number of problems correctly solved is designated by Variable 1. The tendency to use principles is assessed by Variable 2, number of facts uncovered (an inverted index) and Variable 3, number of principles uncovered. In addition, Variable 5 is a weighted combination of the number of problems correctly solved and the relative dependency on principles. Because it assesses jointly proficiency in problem-solving and the preference for prin-

TABLE 2.—PRODUCT-MOMENT CORRELATION COEFFICIENTS FOR SCORES ON BALANCE PROBLEMS TEST AND SUBJECT MATTER ACHIEVEMENT INDICES

Variables	1	2	3	5
6. Mathematics Level	.54**	−.61**	.45**	.71**
7. Average mathematics grade	.28*	−.46**	.34*	.49**
8. Yearly average grade	.28*	−.44**	.29*	.47**
9. Physical Science Achievement	.61**	−.37**	.24	.54**
10. Social Science Achievement	.32*	−.60**	.36*	.61**

* Significant at the five per cent level, but less than the one per cent level.

** Significant at beyond the one per cent level.

ciples, it was thought that it would be more highly related to achievement in mathematics than would either Variables 2 or 3.

The measures of achievement are repeated in Table 2 together with the r's that they were found to have with the above mentioned variables of the BPT.

The only r not significant at the five per cent level or beyond is the value of .24 between Variables 3 and 9. The six correlations significant at the five per cent level or better, but less than the one per cent level, are indicated by an asterisk in Table 2. The remaining thirteen values are all significant at the one per cent level or beyond. These results support the second hypothesis that the number of problems solved and the indices of dependency on principles are positively related to indices of achievement.

The correlations of the weighted score (C + 24 − F) with indices of mathematical achievement are all higher than the analogous r's with Variables, 1, 2, and 3. This gives support to the stated expectancy that the weighted index would be a better predictor of mathematics achievement than any of the other BPT indices.

The degree of association between the level of mathematics training and the weighted BPT score is .71. The BPT problems do not employ any of the so-called 'facts' taught in high school mathematics courses. Solution of the BPT problems does not depend upon the learning of specialized high school mathematical techniques. The application of a principle in the BPT requires no higher mathematics than does the use of facts; both can be applied without any mathematical training beyond elementary school arithmetic.

As shown in Table 2, the r for the weighted score (Variable 5) with an index of achievement is larger in every case except one than the analogous correlations using any one of the other measures on the BPT in lieu of the weighted score. The one exception is the correlation of .61 between Physical Science Achievement and number of BPT problems correctly solved. The analogous r with the weighted score is .54. In other words, a combined index of preference and proficiency in problem-solving is here indicated to be a better predictor of subject matter achievement than either of the indices separately taken. Since the two predictors, combined in this study on an *a priori* basis, have yielded promising results, it would appear worth while to pursue a much more elaborate study to determine the best relative weight to assign to each variable for predictive purposes.

The finding that Physical Science Achievement has lower correlations with the indices of the use of principles, Variables 2 and 3, than with indices of achievement in both mathematics and social science, (see Table 2) led us to examine the content of the tests of achievement. Observationally, the Physical Science test appeared to contain a greater proportion of items designed to test factual information than did the test of Social Science Achievement. The latter test contained more items requiring the use of principles for their solution. It also appears that mathematics grades in high school depend on effective use of principles. These tentative interpretations suggest the hypothesis that indices of the use of principles on the BPT are more highly related to achievement to the extent that the achievement depends more on effective use of principles rather than factual knowledge alone.

Third hypothesis.—In line with our third hypothesis, we were interested in the extent to which BPT scores would compare favorably with tests of general aptitude as predictors of problem-solving achievement, i.e., indicated by achievement in mathematics. Specifically, to what extent can the BPT contribute to the prediction of mathematical achievement?

To investigate this question, it was decided to determine what combination of test scoring variables and achievement scores would yield the best prediction of mathematics grades. The three measures selected to predict grades were weighted BPT scores (C + 24 − F), and Language and Non-language scores of the California Test of Mental Maturity. The intercorrelations with the criterion (mathematical grades) are shown in Table 1. The obtained beta weights were .160, .393, and .216 for BPT (C + 24 − F), Language, and Non-language scores, respectively. The resulting regression equation was:

$$X_1' = .100\,X_2 + .080\,X_3 + .050\,X_4 + 71.74$$

where X_1' = predicted mathematics grade

X_2 = weighted BPT score (C + 24 − F)
X_3 = Language Score
X_4 = Non-language Score

The multiple correlation obtained was:

$$R_{1.234} = .623$$

This multiple correlation can best be interpreted here in terms of R^2, the coefficient of multiple determination. This is the proportion of the variance in X_1 (average mathematical grade) that is dependent upon or predicted by X_2, X_3, and X_4 combined. In this case, R^2 was .3895, indicating that 38.95 per cent of the variance in average mathematics grades was accounted for by the weighted BPT score, the Language and Non-language scores, eliminating from multiple consideration whatever the scores have in common.

Of these three test scores, the results indicate that the BPT is the best single predictor of mathematics grades. At the same time, it should be recognized that the particular score chosen for the BPT (C + 24 − F) was the one which had yielded the highest correlation with average mathematics grades. This point is counteracted somewhat by the prediction made prior to initiating the testing of subjects: that this particular score, with its penalty for use of facts in solving the problems, would yield the best predictions (see description of Variable 5 above).

SIGNIFICANCE OF THE BALANCE PROBLEMS TEST

The Balance Problems Test represents a unique contribution in predicting educational achievement. BPT indices of (a) proficiency in problem-solving, (b) degree of preference for principles over facts in problem-solving, and (c) a weighted combination of these are significantly correlated with achievement in three subject matter areas. The last index appears to be as promising, or even more promising, than either a language or non-language test in the prediction of mathematics grades.

These conclusions have both psychological and educational significance, as well as practical ones. In addition to the predictive possibilities of the BPT, the results indicate that mode of intellectual attack on problems, as represented by individual differences in preference for principles, is a significant variable in problem-solving. If these findings continue to be substantiated, new educational and psychological research problems of determining the variables responsible for these individual differences will present themselves.

SUMMARY AND CONCLUSIONS

The development of a new test—the Balance Problems Test—was undertaken in an attempt to measure the extent to which an individual selects and makes effective use of principles as opposed to factual information in problem-solving. By examining the kinds of information selected, it was hoped to tease out the individuals' characteristic problem-solving approach. Specifically, it was hypothesized that the individual who tended to use general principles would be superior to those depending upon the use of specific facts, in school achievement which involved problem-solving behavior.

Six sets of problems containing four sub-problems each were constructed for use in this study. The problems in each set could be solved either through the application of one general principle applying to the entire set of four problems or through the use of successive, discrete facts. All problems had exactly the same format, but the figures and complexity changed from problem to problem. Each subject was free to select either the 'method of principles' or the 'method of facts' (or both) in solving each problem. The information for each problem was hidden under perforated tabs, each labeled as 'Fact' or 'Principle'; and the subject was free to tear off these tabs to obtain the kind of information he believed would best help him arrive at the solution.

The BPT was administered to thirty-nine high school students in one Illinois community, for whom certain indices of educational achievement and of general aptitude were available. By means of correlational analysis, the following results were secured:

1) Those who take more mathematics courses in high school make more use of principles in the BPT ($r = .71$); they use fewer facts

($r = -.61$), even though the use of principles appears to require more time ($r = .20$).

2) While the use of facts is considered the easier method of solution, those who use facts seem less likely to obtain the correct answer ($r = -.29$).

3) Performance on the BPT, as measured by the weighted score, is a better predictor of mathematics grades than either a Language or a Non-language test of general ability.

4) Physical Science Achievement Test scores are positively correlated to a greater extent with the number of BPT problems correctly solved ($r = .61$) than is achievement on the Social Science Test ($r = .32$). However, scores in Social Science yield a lower correlation with use of BPT facts ($r = -.60$) than the scores in Physical Science ($r = -.37$).

Several distinct contributions appear to follow from our results: First, the techniques used in problem-solving emerge as psychological variables capable of serving as differential predictors of school achievement. Since individuals make characteristic use of these methods (*1*, *2*), the latter could be used in isolating the teachable elements of problem-solving. In terms of educational practice, further investigation into this area could be far reaching.

Second, the tab format of the BPT provides a situation in which the subject is able to select the kinds and amount of information he believes will best enable him to solve the problems. This technique might be extended and applied to situations where the effects of knowledge of results on subsequent test behavior can be measured as well as to obtain measures of selected personality variables (e.g., motivation, rigidity, perseveration). The effects of the ambiguity offered by the more abstract principles might possibly be extended to measure personality.

Third, the tab method might well be applied to situations where the initial selection and use of a specific series of procedures is an important aspect of the individual's subsequent training.

REFERENCES

1) B. S. Bloom and Lois J. Broder. "Problem-solving processes of college students." *Suppl. educ. Monogr.*, University of Chicago Press, 1950, No. 73.

2) E. L. Gaier. "Selected personality variables and the learning process." *Psychol. Monogr.*, 1952, 66, No. 17 (whole No. 349).

3) — — —. "The rôle of knowledge in problem-solving." *Progressive Education*, 1953, 30, 138–141.

NORMATIVE TEST DATA FOR ADULT MALES OBTAINED BY HOUSE-TO-HOUSE TESTING

ELIZABETH P. HAGEN AND ROBERT L. THORNDIKE

Teachers College, Columbia University

As part of a project[1] to study the feasibility of carrying out one or more nation-wide censuses of aptitudes in the adult population, a series of pilot studies was carried out with groups of men tested at their residence units. In several of these pilot studies data were obtained for a set of three tests of ability—a vocabulary test, a mechanical principles test, and an arithmetic reasoning test. In spite of some limitations in representativeness of the samples, these norms for general adult groups appear of sufficient interest to make them available in published form.

The data were obtained during field trials in 1950 and 1951. Most of the field work was done under the direction of the project staff at Teachers College, but one set of surveys was conducted by the United States Bureau of the Census. In all of the surveys the population was defined as males eighteen years old and over. The three brief ability tests (vocabulary, mechanical comprehension, and arithmetic reasoning) were administered by the interviewer to the respondent in his own home. In the surveys conducted by the project staff, the interviewers were primarily college students—either undergraduate upper-classmen or graduate students majoring in psychology. In the surveys conducted by the Bureau of the Census, the interviewers regularly employed by the Bureau for the monthly Current Population Survey were used.

The surveys done by the project staff were carried out in Centre County, Penn.; New York, N. Y.; Springfield, Mass.; and Washington, D. C. Selection of areas was in considerable measure a matter of expediency, locations being chosen in which competent directors could be obtained for the local field work. The sampling procedure, in general, was to break the large, heterogeneous area into a number of segments relatively homogeneous in socio-economic level which were then sampled at random in proportion to the total population of the area. Blocks and dwelling units from

---

[1] The work reported here was done for the Human Resources Research Center, Lackland AF Base, under Contract AF 33(038)-13474.

the selected segments were drawn at random also, so that the resulting sample of units within an area was a 'probability sample'.

The survey by the Bureau of the Census was conducted in seven of the sample areas which serve as sampling units for the Bureau's Current Population Survey. These seven areas were Atlanta, Ga.; Battle Creek, Mich.; Fort Wayne, Ind.; Greenwood, S. C.; Lancaster, Penn.; New York, N. Y.; and Portland, Maine.

Although an attempt was made to locate units in different parts of the country and in different types of communities, the areas which were selected cannot be considered representative of the country as a whole. The most notable omission is the lack of communities west of the Mississippi River and in Southern rural areas. However, the data obtained in the surveys are believed to be of some interest in showing (a) the general level of performance on the tests, (b) the order of magnitude of community differences; and (c) the relationships between education, socio-economic status, and age and performance on the tests.

MATERIALS

Multiple Choice Vocabulary Test.—The test was a twenty-word test made up of two words from the vocabulary subtest of each of ten levels of the IER Intelligence Scale CAVD.[2] It is a steeply graded power test in which almost everyone can get a score, but only a very few get perfect scores. The test was presented to the respondent on a card. The respondent had to read each word and the five choices for himself and give the number of the correct answer. The test was continued until the respondent failed five consecutive items or until he refused to continue. The score was the number of correct responses.

Mechanical Comprehension Test.—This test consisted of the following eighteen items chosen from Form A of the Differential Aptitude Tests, Mechanical Comprehension Test: 1, 4, 9, 12, 17, 20, 25, 28, 33, 36, 41, 44, 49, 52, 57, 60, 65, 68. Items were cut out and mounted on 5 x 8 cards. They were shown to the respondent one at a time, and, except in the case of refusals, each respondent took all eighteen items. In cases where the respondent could not read, the items were read to him. The test was scored number right minus one half the number wrong.[3]

---

[2] Robert L. Thorndike, "Two screening tests of verbal intelligence," *J. Appl. Psychol.*, 1942, 26, 128–135.

TABLE 1.—TEST RELIABILITIES AND INTERCORRELATIONS

Test	Mechanical	Arithmetic	Reliability
Vocabulary	.32	.59	.85
Mechanical Comp.		.47	.59
Arithmetic Reas.			.83

Arithmetic Reasoning Test.—This was a twelve-item test steeply graded in difficulty. The test is reproduced in Appendix A. Ten of the problems were chosen from materials gathered in the course of preparing the IER Intelligence Scale CAVD. Two easier problems were added in an attempt to reduce resistance to the test and the piling up of zero scores. Items were presented one at a time on 5 x 8 cards; subjects could use scratch paper; and the answers obtained were reported to the interviewer. As in the mechanical test, items were read to respondents who were unable to read. The score was the number of items answered correctly.

Table 1 shows split-half reliabilities and intercorrelations of the tests for 1060 cases included in one survey by project personnel.

RESULTS

Tables 2, 3, and 4 give the percentile ranks of the scores on each test and the means and standard deviations on each test for a composite of all of the communities in the surveys done by the Bureau of the Census and a similar composite for the communities surveyed by Teachers College. These tables also show the percentile ranks of the scores for the Teachers College surveys when the sample is weighted and adjusted for age and education according to the 1950 U. S. Bureau of the Census reports for the country as a whole.

On the arithmetic and mechanical comprehension tests, the percentile ranks of the raw scores are closely similar for both the Bureau of the Census and the Teachers College surveys. On the vocabulary test, the percentile ranks are quite similar for the lower half of the distribution but show somewhat less agreement for the upper half of the distribution.

However, the percentile ranks of the weighted sample are quite

---

[3] The courtesy of the Psychological Corporation in permitting the use of this material is gratefully acknowledged.

TABLE 2.—PERCENTILE EQUIVALENTS OF RAW SCORES ON VOCABULARY TEST FOR ALL COMMUNITIES

Raw Score	Percentile Rank		
	Bureau of Census Survey	Teachers College Survey	
		Unweighted	Weighted*
0	1	1	2
1	2	2	6
2	3	4	8
3	4	5	12
4	6	8	16
5	9	12	23
6	14	19	31
7	22	28	41
8	32	37	51
9	45	45	60
10	58	53	67
11	69	61	74
12	78	69	80
13	85	76	85
14	91	82	89
15	95	87	93
16	97	92	96
17	98	95	98
18	99.2	98	99
19	99.6	99	99.7
20	99.9	99.9	
N	597	2292	
$\bar{X}$	9.3	9.7	8.2
SD	3.4	4.2	4.3

* Weighted for age and education.

different from the other two. In all three tests in both the Census surveys and the Teachers College surveys, there appeared to be an under-representation of the lower educational levels and over-representation of the higher educational levels. This was probably due to (a) the biases introduced by refusals to take the test in the over-forty-five age group and (b) inclusion of communities in the study which had a higher level of education than the country as a whole.

Table 5 shows the 10th, 50th, and 90th percentiles for each of the communities on each of the tests. As can be seen from the table,

TABLE 3.—PERCENTILE EQUIVALENTS OF RAW SCORES ON ARITHMETIC TEST FOR ALL COMMUNITIES

Raw Score	Percentile Rank		
	Bureau of Census Survey	Teachers College Survey	
		Unweighted	Weighted*
0	1	1	2
1	3	3	6
2	7	7	12
3	15	14	23
4	27	27	39
5	40	40	53
6	52	50	63
7	65	61	73
8	78	72	82
9	87	82	89
10	94	90	94
11	98	96	98
12	99.7	99	99
N	600	2276	
$\overline{X}$	5.8	5.9	5.1
SD	2.6	2.8	2.8

* Weighted for age and education.

certain communities, notably Springfield, Mass., and Lancaster, Penn., tend to rank high on all of the tests. However, most of the communities tend to perform better on one test than on the other two. Rural areas, such as Centre County, Penn., and Greenwood, S. C., tend to be lower than other communities on the verbal test, but do relatively much better on the mechanical test.

Table 6 shows the results for each test according to socio-economic status. Socio-economic status was determined from the weighted sum of three variables. The three variables used were occupation, house type, and neighborhood rating. These were combined according to procedures described by Warner.[4] An 'A' rating represents highest socio-economic status; a 'D', lowest.

The increase in median score from the lowest socio-economic

---

[4] W. L. Warner, M. Meeker, and K. Eells, *Social Class in America*. Chicago: Science Research Associates, 1949.

TABLE 4.—PERCENTILE EQUIVALENTS OF RAW SCORES ON MECHANICAL COMPREHENSION TEST FOR ALL COMMUNITIES

Raw Score	Percentile Rank		
	Bureau of Census Survey	Teachers College Survey	
		Unweighted	Weighted*
−1½ and below	1	1	1
0	2	2	4
1½	4	4	7
3	8	8	12
4½	13	14	19
6	22	24	30
7½	36	36	44
9	52	50	58
10½	66	64	71
12	78	76	82
13½	88	88	91
15	94	92	96
16½	97	97	98
18	99	99	99
N	610	2278	
$\overline{X}$	8.8	8.9	8.1
SD	3.9	4.1	4.2

* Weighted for age and education.

group to the highest was about 1.8 standard deviation units (based on the standard deviation of the total group) in the case of the vocabulary test, 1.4 standard deviation units for the arithmetic test, and 1.0 standard deviation units for the mechanical comprehension test. In each case, there was a regular progression in score from level to level, each rise in socio-economic level being accompanied by a rise in score. It is perhaps worth noting that for the vocabulary test the level A-level B difference is the largest one, while for the other two tests it is the smallest one. That is, the highest socio-economic group is particularly characterized by a high verbal level. These findings for adult groups are consistent with results from many studies of the relationship between father's occupation or family socio-economic status and the scores made by children on intelligence tests.

Table 7 shows the distribution of scores by educational level for

TABLE 5.—SELECTED PERCENTILE SCORES ON THE VOCABULARY TEST, MECHANICAL COMPREHENSION TEST, AND ARITHMETIC TEST FOR COMMUNITIES IN THE TEACHERS COLLEGE AND BUREAU OF THE CENSUS SURVEYS

	Vocabulary			Mech. Comp.			Arithmetic		
	P_{10}	P_{50}	P_{90}	P_{10}	P_{50}	P_{90}	P_{10}	P_{50}	P_{90}
Teachers College Surveys									
Springfield, Mass.	6.8	11.7	16.2	4.7	10.4	15.9	3.5	7.2	10.4
New York, N. Y.	5.9	11.4	16.4	3.1	8.4	13.7	3.0	6.6	10.0
Washington, D. C.	3.7	9.2	15.0	2.4	8.6	14.4	1.9	5.4	10.2
Centre Co., Pa.	3.9	7.4	12.3	4.5	9.6	14.2	2.4	5.4	9.1
Bureau of the Census Surveys									
Atlanta, Ga.	3.4	8.9	13.4	2.9	8.2	14.3	2.7	5.8	9.1
Battle Creek, Mich.	5.1	8.7	12.6	4.1	9.3	14.4	2.6	5.2	8.9
Fort Wayne, Ind.	6.0	9.7	13.7	5.3	9.2	13.7	2.7	6.3	9.4
Greenwood, S. C.	2.9	8.2	12.9	4.1	8.6	13.5	1.7	4.2	8.4
Lancaster, Pa.	6.8	9.6	14.1	5.4	9.1	13.8	3.9	6.8	9.7
New York, N. Y.	5.0	11.5	15.0	2.4	7.7	13.6	2.7	5.9	8.4
Portland, Maine	5.2	9.3	12.2	3.1	8.2	13.7	2.1	6.1	10.0

TABLE 6.—SELECTED PERCENTILE SCORES ON ALL THREE ABILITY TESTS FOR ALL COMMUNITIES IN TEACHERS COLLEGE SURVEYS, BY SOCIO-ECONOMIC STATUS

Percentile and Test	Socio-Economic Statue			
	D	C	B	A
Vocabulary				
10th percentile	1.9	5.2	6.7	9.1
25th percentile	4.5	6.7	9.1	12.0
50th percentile	6.8	9.2	11.9	14.5
75th percentile	9.2	11.9	14.5	16.4
90th percentile	11.4	14.4	16.4	17.9
Mechanical Comprehension				
10th percentile	1.8	3.5	5.1	6.1
25th percentile	4.7	7.7	7.6	8.6
50th percentile	7.3	8.9	10.3	11.4
75th percentile	10.1	11.5	12.8	14.2
90th percentile	12.4	15.3	15.5	16.3
Arithmetic				
10th percentile	1.3	2.7	4.1	5.1
25th percentile	2.6	3.9	5.7	6.8
50th percentile	3.9	5.7	7.6	8.8
75th percentile	5.8	7.8	9.3	10.4
90th percentile	7.6	9.3	10.6	11.3

TABLE 7.—SELECTED PERCENTILE SCORES ON ALL THREE ABILITY TESTS FOR ALL COMMUNITIES IN TEACHERS COLLEGE SURVEYS, BY YEARS OF EDUCATION COMPLETED

Percentile & Test	Last School Grade Completed									
	1–4	5–6	7	8	9	10–11	12	13–15	16	Over 16
Vocabulary										
10th percentile	0.3	0.8	0.9	3.1	4.0	4.9	6.0	7.4	9.2	10.2
25th percentile	0.9	2.6	2.8	5.0	5.8	6.3	7.5	9.4	11.8	12.9
50th percentile	3.3	5.1	5.5	6.4	7.4	8.3	9.7	11.5	13.9	15.1
75th percentile	4.7	7.0	7.9	8.8	10.0	11.1	12.0	13.9	15.7	16.9
90th percentile	7.8	9.2	9.4	11.2	12.3	13.0	14.4	15.8	17.0	18.2
Mechanical Comp.										
10th percentile	−1.8	−0.95	1.4	2.5	2.4	2.7	4.3	5.0	5.1	7.0
25th percentile	2.9	3.4	4.2	4.4	5.7	5.4	6.9	7.0	7.9	8.9
50th percentile	6.6	6.0	6.5	7.6	8.2	8.0	9.4	10.1	10.2	11.6
75th percentile	9.4	8.9	9.2	11.8	10.7	10.7	11.8	12.5	14.5	13.9
90th percentile	11.1	12.4	11.8	12.3	12.6	13.0	19.2	15.1	16.4	16.3
Arithmetic										
10th percentile	0.5	0.0	0.8	1.6	2.1	2.6	3.3	4.0	5.0	5.8
25th percentile	1.3	1.6	2.1	2.8	3.2	3.7	4.5	5.3	7.2	7.3
50th percentile	2.8	2.8	3.2	3.9	4.4	4.9	6.3	7.4	8.8	9.0
75th percentile	3.8	4.2	4.6	5.5	6.6	7.0	8.3	9.1	10.3	10.4
90th percentile	4.7	5.0	6.0	7.1	8.1	8.4	9.4	10.3	11.2	11.4

each of the three tests. The vocabulary and arithmetic tests both show steep and steady increases in median score as the level of education increases. The mechanical comprehension test shows an increase in median score as educational level increases, but the increase is neither as steady nor as large as that for the other two tests. On the vocabulary test, the increase in median score from the lowest to the highest educational level is equal to 2.8 standard deviations; on the arithmetic reasoning test it is equal to 2.2 standard deviations, and on the mechanical test to 1.2 standard deviations. These relationships between education and scores on the tests—especially the vocabulary and arithmetic tests—are consistent with the correlation of .70 which was reported by Tuddenham for the relationship between education and AGCT scores.[5]

Table 8 shows the distribution of scores for the three tests for

---

[5] Read D. Tuddenham, "Soldier intelligence in World Wars I and II," *The American Psychologist*, 1948, 3, 54–56.

TABLE 8.—SELECTED PERCENTILE SCORES ON ALL THREE ABILITY TESTS FOR ALL COMMUNITIES IN TEACHERS COLLEGE SURVEYS BY WEIGHTED AGE GROUPS*

Percentiles and Tests	Age Groups			
	18–25	26–35	36–45	Over 45
Vocabulary				
10th percentile	2.7	2.4	2.5	1.3
25th percentile	4.7	5.2	5.1	4.3
50th percentile	6.7	7.9	7.8	7.3
75th percentile	9.4	10.8	11.4	11.0
90th percentile	11.5	13.6	14.0	14.3
Mechanical Comprehension				
10th percentile	0.8	2.2	2.3	2.7
25th percentile	5.0	5.2	5.4	5.3
50th percentile	7.9	8.3	8.2	7.6
75th percentile	10.7	10.8	10.8	10.9
90th percentile	13.5	13.2	13.4	13.2
Arithmetic				
10th percentile	0.5	1.4	1.3	1.3
25th percentile	2.4	2.8	2.7	2.6
50th percentile	4.0	4.4	4.3	4.1
75th percentile	6.5	6.8	6.9	6.5
90th percentile	8.5	8.6	8.7	8.5

Age groups weighted for educational level according to 1950 Bureau of the Census data on education for the country as a whole.

age groups which have been weighted according to 1950 Bureau of the Census data on educational level for different age groups for the country as a whole. The weighting was made to correct for educational biases in the sample. In the sample, the higher educational levels in the eighteen to twenty-five year old group were under-represented, presumably because a large number of these men were away from home at the time of the survey and were not available for interviews. In the older age groups, the lower educational levels are under-represented, in part, at least, because the older men with less education tended to refuse to take the tests.

Before the age groups were weighted for educational level, the increase in test scores by age followed the pattern for education. After the correction, the median scores on all three tests rise slightly in the twenty-six to thirty-five year old group, then drop off in the thirty-six to forty-five and over-forty-five age groups. In the vocabulary and arithmetic tests, the median score for the upper two age groups is slightly above that for the eighteen to twenty-five age group. In the mechanical comprehension test, the median score for the over-forty-five age group is slightly below that for the eighteen to twenty-five year old group. However, all of the differences in median scores are small—the largest being approximately .3 of a standard deviation unit.

APPENDIX A

Arithmetic Reasoning Test

1) If 2 pencils cost 5¢, what will 4 pencils cost?

2) You buy a newspaper for 5¢ and a pack of cigarettes for 21¢. If you give the clerk 50¢, how much change should you get?

3) What does a pound of candy cost when you pay 10¢ for ¼ of a pound?

4) A man bought land for $400. He sold it for $445, gaining $15 an acre. How many acres were there?

5) How much is ⅞ of 28?

6) 20 = how many times 12?

7) How many times as long as 8 feet is 12 yards?

8) Counting that 25 dozen sheets of paper are worth 10¢, how many sheets of paper are worth ⅕ of a cent?

9) How many quarters of a quarter is half of a half?

10) How long will enough food for 400 men for 120 days last 250 men?

11) A man started with $12,500 and doubled his capital every year for five years. How much had he at the end of the fifth year?

12) A salesman was to receive 10¢ for every dollar's worth of merchandise he sold, an extra 10¢ for every dollar over 100, and a third 10¢ for every dollar over 1000. The salesman sold P plus $1000 worth of merchandise. How much should he receive?

PREDICTING ATTRITION-SURVIVAL IN FIRST YEAR ENGINEERING

JOHN P. MALLOY, BOLESLAW WYSOCKI, AND LEO F. GRAHAM

Marquette University

During recent years Marquette University engineering students have been administered the Minnesota Paper Form Board Form AA, as part of a battery of entrance tests. The American Council on Education, 1949 edition, the Coöperative Mathematics Test Form Y, and the Coöperative Reading Test Form RX, also have been included in the battery. There has been some question about the effectiveness of the Minnesota Paper Form Board. Some advisors have felt that the intercorrelations between the Minnesota Paper Form Board and other tests in the battery, particularly the ACE-Q score, were such that relatively little was gained by its inclusion. This study was undertaken in order to evaluate the importance of the Minnesota Paper Form Board as a predictor of success in engineering college.

A second purpose of this study was to determine the effectiveness of each test for predicting survival in Marquette's Engineering College. The majority of the studies investigating the prediction of college success by means of psychological measures have utilized regular correlation methods with college grades as the criterion. Since survival in the engineering curriculum ordinarily implies entrance in the profession, it was felt that survival was an appropriate and logical criterion. Although survival for the full four years of work would be a better criterion, survival for the first year was used because of the unavailability of longitudinal data encompassing a longer period of time.

Minnesota Paper Form Board scores, A.C.E. scores, and Coöperative Mathematics scores were available for one hundred and ninety-four students who had enrolled in Marquette Engineering College in the fall of 1952. The sample represented seventy-one per cent of the class. Although the reading test scores were available, they were not utilized as part of this study. The MPB scores were unavailable for the remaining twenty-nine per cent of the class and hence this group could not be included in the study.

Of the one hundred and ninety-four students, one hundred and sixty-one survived the first year. Causes for drop-out were disregarded and no distinction was made betweeen students that were dropped by the college and those who dropped of their own accord for various reasons.

The attrition rate in this sample was eighteen per cent. An examination of the entire class revealed that thirty-one per cent of the students failed to remain the entire year. The difference between the sample's attrition rate and that for the entire class probably can be traced in part to ramifications of Marquette's admission policy.

The university regularly accepts high school graduates who rank in the top two-thirds of their class without further qualifications. If a student cannot meet this standard, it is possible for him to gain entrance provided that he performs satisfactorily on the freshman test battery. This battery includes the tests previously mentioned except for the Minnesota Paper Form Board. It also includes the Coöperative Mechanics of Expression Test. All students whose scores were used in this study were those admitted on the basis of satisfactory high school performance. For this reason they were a select group.

From the differences of the rate of attrition between those students admitted on the basis of high school rank and those who failed to meet the standard but are later admitted on the basis of their performance on the psychological tests, it would seem that the latter group have a much poorer chance for success than those that are admitted without qualification. At the present time the admission policies of the University are undergoing scrutiny as part of a self-survey now being conducted throughout the University by the administration and by the faculty.

TREATMENT OF THE DATA

The usual method for determining the predictive effectiveness of psychological measures has been the utilization of correlation methods where both independent and dependent variables are numerical variables. Since the attrition-survival criterion results in a dichotomy, the usual regression techniques were inappropriate for this study and another design was necessary.

Discriminant analysis and the serial correlation which can result from it satisfy this need. Fisher (*1*) developed the former technique which basically involves ascertaining appropriate weights for a series of variables yielding maximum separation in two groups, each

TABLE I.—MEAN SCORES FOR THE SURVIVAL AND ATTRITION GROUPS

Numerical Variable	Survival N-161	Attrition N-33	Difference
ACE Q (x_1)	53.31	47.12	6.19
ACE L (x_2)	67.27	61.45	5.82
MPB (x_3)	48.36	42.42	5.84
Math (x_4)	25.48	21.88	3.60

of which is assumed to be normally distributed. Wert, Neidt and Ahmann (*2*) have illustrated the use of this design for psychological data. Accordingly the data in this study were analyzed with discriminant analysis except where the computation of Pearson Product Moment r's was appropriate.

RESULTS

An inspection of Table I provides a review of the performance of the attrition-survival groups. The survival groups performance surpasses the attrition group in each instance. The solution of the coefficients provided the function:

$$V = .00026832x_1 + .00006118x_2 + .00027178x_3 + .00003033x_4$$

where

V = deviation relative score

x_1 = deviation ACE-Q score

x_2 = deviation ACE-L score

x_3 = deviation MPB score

x_4 = deviation Math score

TABLE II.—COEFFICIENTS OF CORRELATION BETWEEN VARIABLES

N = 194

Survival R_{bis}	$Q_{(x_1)}$	$L_{(x_2)}$	$MPB_{(x_3)}$	$Math_{(x_4)}$
	.398	.239	.375	.258
Q (r)		.408	.386	.444
L (r)			.163	.371
MPB (r)				.341

$R_{bis(x_1x_2x_3x_4)} = .473$

$R_{bis(x_1x_2x_3)} = .472$

$R_{bis(x_2x_3x_4)} = .424$

$R_{bis(x_1x_2x_4)} = .413$

$R_{bis(x_1x_3x_4)} = .446$

$R_{bis(x_1x_2)} = .407$

$R_{bis(x_1x_3)} = .465$

$R_{bis(x_1x_4)} = .408$

$R_{bis(x_2x_3)} = .416$

$R_{bis(x_2x_4)} = .301$

$R_{bis(x_3x_4)} = .399$

TABLE III.—TESTS OF SIGNIFICANCE RESULTING WHEN ONE TEST WAS DELETED FROM A COMBINATION OF TESTS PREDICTING THE CRITERION

Predictor Combination (R_{bis})	Test Eliminated	F	d.f
$x_1x_2x_3x_4$	x_4	.1	1,189
$x_1x_2x_3x_4$	x_3	†5.6*	1,189
$x_1x_2x_3x_4$	x_2	2.4	1,189
$x_1x_2x_3x_4$	x_1	4.2*	1,189
$x_1x_2x_3$	x_3	5.5*	1,190
$x_1x_2x_3$	x_2	.7	1,190
x_1x_3	x_3	5.6*	1,191

* Significant at the five per cent level.

† This is interpreted to mean that a significant loss in prediction effectiveness results at five per cent confidence level when this test is eliminated from the battery.

From a review of Table II the predictive effectiveness of each variable may be observed. The zero order Pearson r's between each pair of numerical variables were also computed. A test of the significance of the difference from zero of the smallest biserial r, that between L and survival, was made. It was highly significant. The smallest Pearson zero order r, that between the ACE-L score and the MPB, was significant at the five per cent level but not at the one per cent level.

To determine the best predictive combination the discriminant functions were computed in such a fashion as successively to delete each of the prediction variables. The resulting F-tests are recorded in Table III. The L-score and the Coöperative Mathematics Test may be eliminated from the prediction equation since neither contributes significantly. It is obvious that the Q score and the MPB score make the best two variable predictors.

DISCUSSION

A surprising outcome of this study was the ineffectiveness of the mathematics score in the prediction of attrition-survival. It is possible that an extension of this study to include students dropping out in the sophomore year might reveal an increase in the importance of this variable. It is well known that the calculus is a major stumbling block for embryonic engineers.

The importance of the Minnesota Paper Form Board may result from the inclusion of engineering drawing and descriptive geometry in the freshman curriculum. It seems certain that it should be retained in the test battery for engineering students.

As an outcome of this study a table with a two-way entry was constructed, using the ACE-Q score and the MPB score. This table was fashioned so that it can be entered with the percentile scores of both the ACE-Q and MPB scores, and the chances in one hundred of a student surviving is determined. When one enters this table with the Q and MPB scores at the fifth percentile, a student's chance of surviving is sixty-one in 100. If a student achieves scores at the 95th percentile on both measures, his chances for survival for the first year of engineering at Marquette University are ninety in one hundred.

CONCLUSIONS

1) The ACE-Q score and the Minnesota Paper Form Board score were the best predictors of survival in first year engineering at Marquette University.

2) These two measures are relatively effective as predictors of attrition-survival.

3) The Coöperative Mathematics Test contributed least to prediction of the criterion.

4) Since a legitimate question can be raised as regards the representativeness of the sample when compared with the entire class, caution must be used in generalizing from this study.

SUMMARY

A study of attrition-survival was made at the Marquette College of Engineering. By means of discriminant analysis and serial correlation it was discovered that the ACE-Q score and the Minnesota Paper Form Board were comparatively effective as predictors of survival. A multiple biserial r equal to .473 resulted when four predictor variables were used. When the two best cited tests were used, a serial correlation coefficient of .465 was obtained.

REFERENCES

1) R. A. Fisher. "The use of multiple measurements in taxonomic problems," *Annals of Eugenics*, 7: 179–188, 1936.

2) J. E. Wert, C. O. Neidt, and J. S. Ahmann. *Statistical Methods in Educational and Psychological Research.* New York: Appleton-Century-Crofts, Inc., 1953.

BILINGUALISM AND BICULTURALISM

JAMES P. SOFFIETTI

Syracuse University

The literature on bilingualism (*1*, *2*, *4*, *5*, *6*) is voluminous. It deals chiefly with its effects on the mental and linguistic development of the child and his school achievement, as well as his personal and social adjustment. As a corollary, bilingualism also raises the question as to the advisability of teaching a second language in our schools and that of the time and method of such instruction.

These problems have been discussed and investigated from many points of view both in this country and abroad. On one sole point there seems to be general agreement: that the findings so far are inconclusive and that the problem needs further careful investigation. The general implication seems to be that bilingualism is a much more complicated condition of affairs than that of the use of two languages by an individual.

If bilingualism is not a simple concept, then it needs further differentiation and clarification. From the point of view of dictionary definitions, the concept of bilingualism is quite simple. The American College Dictionary defines it (1) as 'the habitual use of two languages'; and (2) 'the ability of being bilingual'. It defines the bilingual as a person 'able to speak one's native language and another with approximately equal facility'.

The complications begin to arise when one attempts to apply the dictionary definition to specific cases. It then loses most of its value as a working concept. One soon discovers that most of the practical situations referred to as 'bilingual' involve factors that extend far beyond those of the 'habitual use of two languages'. Investigators have usually been aware of this, but somehow have not realized the necessity of isolating such factors from the basic concept of bilingualism.

Thus we find the researcher talking about 'bilingualism' when he is referring not only to the use of two languages, but also to the fact (and at times mainly so) that the individual or group under investigation is living in a different 'cultural atmosphere' at home or in his neighborhood from the one prevailing in the community at large. In other words, to the investigator bilingualism refers not only to two distinct patterns of 'linguistic habits', but also to distinct patterns of 'cultural habits' in all of their anthropological meaning. His problem is not merely one of bilingualism, but also one of what we may justifiably call 'Biculturalism'.

Many students of bilingualism appear to assume that the concept of language already comprises the factors of cultural differences. This is a misconception. While it is true that language behavior is an integral part of cultural behavior and that the latter would be impossible without its linguistic components, it does not mean that to a given language system there has to correspond a specific cultural structure or area. It is possible for otherwise totally unrelated cultures to share in one linguistic system and for totally unrelated languages to share in one cultural system. Language, race, and culture are not necessarily correlated (*9*). Even the fact that language tends to canalize the speaker's response does not mean that it is inseparably bound to a specific culture (*8*).

By keeping a clearcut distinction between the concepts of 'bilingualism' and 'biculturalism', the investigator will find it easier to outline his problem of research, conduct his investigation, categorize his findings and report the facts; and the classroom teacher will find it easier to solve his 'bilingual' problems. But before discussing these two concepts in greater detail, there is another important one that should be isolated from the current meaning of the term bilingualism.

Researchers have also been aware of the fact that there are degrees of bilingualism. For various reasons, not all bilinguals are proficient to the same extent in the two languages used. In such cases we usually speak of 'linguistic accents'. They are due to various forms of interference between the patterned habits of the mother tongue and those of the second language with results of distortion, substitution, or omission, of speech sounds or grammatical structures, or both.

To the same extent we may speak of 'Cultural Accents'. This term is not intended to be a metaphorical expression. 'Cultural accents' are real and of the same nature as the linguistic ones. They, too, are the result of conflicting habit patterns persistently interfering with each other and resulting in any number of distortions or outright substitutions. They may be detected at any point and in various degrees throughout the daily routine activities of a non-native member of a society: from table manners to non-verbal forms of greeting; from gait to ways of driving a car; from choice of foods to manner of preparing them; from little things about dressing to features in house furnishing; from attitudes towards members of the family to attitudes towards one's 'superiors' in business, and in intellectual or socio-economic status. They can be perceived in non-verbal expressions of emotion as well as in the value system of the individual under consideration.

At times, however, a cultural accent may be so subtle that all we can say is that there is something 'foreign' about the person in question without our being able to put our finger on the distinguishing feature involved.

Learning to behave in conformity with the non-verbal cultural patterns of another society can be as much of a task as learning its verbal patterns, i.e., its language. To get rid of cultural accents can be just as difficult as getting rid of linguistic ones: both are based on deep-rooted systems of habit, some of which have strong emotional basis.

The concepts of linguistic and cultural accents can play an important rôle in the definition of certain types of minority groups. In many instances the only distinguishing feature of a minority is the presence of a certain degree or type of cultural or linguistic accents, or both, which have survived as vestiges through any number of generations, while conformity exists in all the other aspects of their societal behavior.

By keeping a clearcut distinction among the concepts that have been developed above, instead of gathering them under the single label of bilingualism, the research worker will be able to distinguish between the following four basic types of situations:

(1) bicultural-bilingual
(2) bicultural-monolingual
(3) monocultural-bilingual
(4) monocultural-monolingual

We are omitting from this discussion the possibility of multilingual and multicultural situations, as well as further differentiations within the bicultural and bilingual concepts as suggested, e.g., by Mary R. Haas and others. (*7*).

(1) The bicultural-bilingual person. He is the one that is usually considered as the true 'bilingual'. He is able, or actually does, participate intimately in two cultures (*3*). An example would be the child of immigrant parents who at home continue to speak the

native tongue and maintain most of the native customs, beliefs and value patterns.

(2) A bicultural-monolingual case would be that of a child of immigrant parents who, for a number of reasons, have given up speaking their native tongue or whose native language happens to be English (such as a British or Irish family), but who have not given up their native customs, beliefs or value patterns.

(3) A monocultural-bilingual example is that of a child who participates in one sole culture, but who has learned to use a second language either in school or at home. This would be the case, e.g., when one or both of the parents are bilingual and have decided that at least one of them shall always speak to the child in the foreign language, although the remaining aspects of cultural behavior are those of the community at large.

(4) It goes without saying, that the most common situation in the United States is the monocultural-monolingual one. In this case, the child does not have to contend with linguistic and cultural conflicts. (We are omitting from this paper the discussion of problems of sub-cultural nature.)

In the first three types of situations, the experienced investigator would detect more often than not various degrees of foreign accent in either the language or cultural patterns, or both. Perhaps cases of perfect bilingualism and biculturalism are extremely rare, if at all possible.

If one studies the literature dealing with the effects of bilingualism on the intellectual and social development of the child in the light of the above new concepts, he will find that most of the difficulties and retardations indiscriminately ascribed to bilingualism are rather due to the bicultural aspects of the situation under consideration. It is the living in two distinct cultures either overtly or in one's internal life that might create adjustmental problems. It is a conflict between ways of life, beliefs, customs, value systems and not necessarily one between language systems.

A person learning a second language in a monocultural setting will not automatically learn a whole new set of cultural patterns and develop cultural conflicts or accents. As a matter of fact, it is seldom that a person studying a second language in our schools ever gets beyond the 'synonym stage' (i.e., beyond the purely bilingual stage as defined in this paper) of foreign language learning: all he usually learns is another way of referring to the same or similar things, situations, or events. 'Mère' is just another word for 'mother'; 'je vais chez moi' another way of saying 'I am going home'. Rarely do his meanings extend beyond the simple denotation to include emotional and cognitive associations common to all the participants in the culture where that particular language is spoken. It is only by living in the culture where the language is spoken that the full meanings of its utterances are acquired. Learning French or German, or Urdu, in a monocultural setting will no more create problems of intellectual and social nature than learning Esperanto.[1]

In conclusion, it is hoped that the concepts of 'biculturalism' and 'cultural accents' developed in this article will aid the educator, the psychologist and the sociologist in their effort to isolate the distinctive elements that characterize the so-called bilingual person or situation. It is also hoped that these concepts will help the bicultural person to a better understanding of himself.

REFERENCES

1) Seth Arsenian. "Bilingualism in the postwar world." *The Psychological Bulletin*, 1945 (February), pp. 65–86.

2) — —. *Bilingualism and Mental Development.* New York: Teachers College, Columbia University, 1937.

3) J. S. H. Bossard. "The bilingual as a person—linguistic identification with status." *American Sociological Review*, 1945, 10: 699–709.

4) Einar Haugen. *A Bibliography of Bilingualism.* Madison: University of Wisconsin. (in preparation.)

5) T. Manuel Herschel. "Bilingualism." *Encyclopedia of Educational Research*, 1950 (rev. ed.)

6) Werner F. Leopold. *Bibliography of Child Language.* Evanston: Northwestern University Press, 1952, p. 115.

---

[1] The above is in no way meant to minimize the cultural value of studying a foreign language. It will generally be agreed, however, that these values can only be gotten when the foreign language class is enriched by illustrated and dramatized discussions about the foreign culture and by comparisons of important aspects of it and its language with corresponding features in ours. As every anthropologist well knows, without some knowledge of the language it is impossible to gain a thorough understanding of a culture for the simple reason that many of the cultural traits depend on the language components for their very existence and configuration. How could anyone on his own even begin to understand our way of life without having some knowledge of English?

7) Claude Levi-Strauss, et al. "Results of the conference of anthropologists and linguists." *Memoir No. 8, I.J.A.L.*, 1953 (April), p. 42.

8) Eric H. Lenneberg. "Cognition and Ethnolinguistics." *Language*, 1953, 29: 463–71.

9) Edward Sapir. *Language*. New York: Harcourt, Brace and Co., 1921, p. 221 ff.

WAR-TIME ACCELERATES TEN YEARS AFTER

MARIE A. FLESHER and S. L. PRESSEY

The Ohio State University

During World War II acceleration was the most hotly debated of all educational topics, and most colleges and technical schools had some form of accelerated program. When the war ended, most of these programs seemed quickly to have been dropped, and most educators seemed desirous to forget the topic as quickly as possible. However, the Korean war and likely indefinite continuance of the draft or related measures in view of continuing international tensions, makes the matter again of at least potential importance. Moreover, the few studies which have been made of war-time programs of acceleration have shown not the anticipated unfortunate effects, but rather the reverse! Thus the writers studied intensively the one hundred and four women who in the four colleges of Arts, Education, Commerce and Agriculture of The Ohio State University completed a four-year undergraduate program in three years or less, from 1941 through 1945. (*4*) The average academic record of these one hundred and four accelerates was found to be better than the records of one hundred and four other women graduating in the same period who entered at the same age, made approximately the same scores on the test of general ability given at entrance, and took the same type of program, but who spent the usual four years (three years and nine months, or four years if graduation was in August) in obtaining an undergraduate degree. The accelerates also participated reasonably in extra-curricular activities. And a follow-up a couple of years after graduation showed them beginning their careers in as promising fashion as the paired 'regular' cases.

Terman's notable continuing studies of his gifted child group have added impressive new evidence of the value of rapid progress in school of able young people. (*5, 6*) The Ford Foundation projects have reëmphasized the values of acceleration at the college level! Under all these circumstances, a further follow-up of the writers' war-time accelerate cases seemed desirable, to see how they were doing approximately ten years after graduation, and how they then regarded their accelerated programs. This paper is a brief summary of such a study.

CASES AND METHOD

When the above mentioned one hundred and four women accelerates were first being studied, it was hoped that a follow-up might be made. Hence their names, as well as the names of students paired with them but taking the usual three years nine months or four years to complete an undergraduate program, were saved. In 1946, forty-one more young women completed an undergraduate degree in three years or less, they were 'paired' with forty-one other women students. The present study is concerned with the total one hundred and forty-five accelerates and their paired cases who graduated from 1941 through 1946 and is based upon data gathered in the spring of 1954. The subjects at that time averaged around thirty-two years of age and had graduated an average of ten years previously.

For sending to the total of two hundred and ninety alumnae, a simple and straightforward two-page questionnaire was prepared, which could be very quickly filled in, almost entirely by checking. It was mimeographed on University stationery and began with the following statement:

> Dear Ohio State University Alumna:
>
> You were a student at O.S.U. during the Second World War, when certain adjustments of program were made. Campus life was in many respects different from normal times, and some students worked in war industry. An appraisal of such war-time adjustments and conditions as seen in retrospect seems desirable, especially since the cold war and possible crises might again bring educational changes. Your answers to the following questions can be very helpful; most require only a check after the answers you think correct.

Then followed in order questions about employment while in school, acceleration, extra-curricular activities, further schooling after the undergraduate degree, marital status, employment, and participation in community affairs. Special effort was made to phrase the entire document so that the recipients would not know that acceleration was a special matter of concern. And indeed the aim was, as the introductory statement said, to obtain broad appraisals regarding student life at the University in the war period.

TABLE 1.—PERCENTAGES OF 112 ACCELERATES AND 100 REGULARS WHO, DURING THE PERIOD WHEN THEY WERE UNDERGRADUATES, WERE EVER EMPLOYED

	Acc.	Reg.
Were at some time employed	79	86
employed while in school	73	68
during vacations or quarters not in school	31	65
full time	12	30
part time	69	73
Percentages of those employed who felt that employment		
put a strain on health	3	7
interfered with schooling	7	8
unfortunately limited social life	19	18
was a good experience	93	87

That slightly fewer of the accelerates did so might be the result of not only more unremitting school-going or heavier loads but also less need, since completing a degree in three years obviously costs less than taking a year more. And, since accelerates and their controls began school at the same age, the accelerates finished a year younger.

Yet more interesting are the last four lines of the table. Only rarely did either group consider that employment put a strain on health, and slightly fewer of the accelerates than of the regulars so thought. Neither did either group feel that employment interfered with schooling. Only a few thought that it unfortunately limited social life, with no significant difference between the two groups. The great majority—and slightly more of the accelerates—thought that employment during the school years was a good experience.

The first part of Table 2 shows the percentages of the accelerate and the regular groups who ever took a four-quarter schedule, obtained credit by examination, took extra heavy course loads, or tried any combination of the above three methods of expediting progress through a college program. A few of the students who took four years for a degree did at some time take one of these steps; but they then were out of school, were ill, or for some reason did not actually shorten their over-all time in college. On the other hand, ninety-two per cent of the accelerate group, at least once, went four quarters in school; over half took extra heavy schedules; twenty-nine per cent got some credit by examination; and over

One of the total two hundred and ninety students was found to have died, and persisting search failed to locate the addresses of ten others. To the remaining one hundred and thirty-nine accelerates, and one hundred and forty students who took the regular time, the inquiry forms were mailed. Since the post office did not return them as undelivered it is assumed that the addressees received them, though a few may have very likely been accepted by a member of the family, or otherwise not come to the attention of the addressees. With each inquiry form was included a stamped addressed envelope for return.

The forms were mailed out April 7, 1954, and a follow-up was sent to those from whom replies had not then been received, on June 11. By August 1, eighty-one per cent (one hundred and twelve cases) of the accelerates had returned filled-in forms, and so had seventy-one per cent or one hundred cases of the regulars. A greater return from the accelerates might be attributed to greater interest on their part because they were aware of having adjusted more to the war-time situation. A further finding can perhaps be explained similarly. At the end of the questionnaire was the statement: "If you would like a report of the results of this study, give your name and address. But if you prefer, return the form unsigned." Eighty per cent of the accelerates, but only sixty-eight per cent of the regulars gave their names and addresses.

Though seventy-six per cent return on a questionnaire (for the two groups together) might be considered good, the question is as always as to whether selective factors may seriously have biased the results. Replies ranged from the highly favorable to the strongly and anonymously critical; it is believed that they are reasonably representative of the total group solicited. And the great majority of replies seem to have been carefully considered; additional written commentaries often extended to the backs of the blanks. In general, these alumnae seemed pleased to have had their opinions asked, and were interested in the topics.

RESULTS

The first question asked was "During the period when you were an undergraduate, were you ever employed? If so, indicate when. While in school? —" and so on. The Table I summarizes the answers. It shows that the great majority found it possible to earn some money during the years when they were going to college.

TABLE 2.—EXPERIENCE WITH, AND JUDGMENTS OF, 'ACCELERATION'

	Acc.	Reg.
Percentages of each group who ever		
took a four-quarter yearly schedule	92	11
obtained credit by examination	29	12
took extra-heavy course schedules	57	15
used two or more of above methods	57	6
Percentages who felt that accelerated programs		
were a strain on health*	7	15
prevented best school work*	12	15
undesirably limited social life*	9	16
were a desirable challenge	62	40
desirably saved time	79	34

* Included here are a few judgments that undesirable results occurred at times though not generally, and a few replies of doubtful or 'not sure'.

half used two or more of such methods. In the judgment of the writers, there should be more use of credit by examination; and four-quarter schedules should then become less necessary. (*4*, *7*)

Even though there may have been too continuous school-going and too much heavy course work, the second part of the table indicates that very few of the accelerate group thought accelerated programs a strain on health, not many judged that acceleration "prevented the student from doing as good work as he normally would," or that accelerated programs "undesirably limited social life." Instead, those not accelerating thought, more frequently, that such undesirable outcomes followed. And two-thirds of the accelerate group thought that acceleration presented a desirable challenge, while over three-fourths felt that desirable time was saved. Apparently few of the accelerates looked back on their college programs some ten years or so earlier with any feeling that damage had so been done them; rather, most felt that their acceleration had brought gains.

It has been held that accelerated programs cause too unremitting work and interfere with student activities, these being evaluated as making major contributions to a student's real growth in the college period. The writers' studies during and immediately after the war showed the accelerates only slightly less active than students taking the regular four years for an undergraduate degree. But it was felt desirable to get the reaction of these alumnae on

TABLE 3.—PERCENTAGES IN EACH GROUP WHO ENGAGED IN ACTIVITIES

	Acc.	Reg.
Activities related to War, as Red Cross	65	65
three or more such activities	9	13
Usual student activities	76	88
three or more such activities	42	60
Percentages listing two or more who felt that activities		
took too much time and energy	10	15
were sometimes too frivolous	18	18
were of value in broadening experience	86	87

these matters, as they saw them ten years after. Table 3 summarizes the results. About as many of the accelerates engaged in activities related to the war as did the regulars. But more of the regular students went into the usual student activities—doubtless because they had more time, and partly probably because student activities are organized in terms of the conventional four-year autumn-winter-spring schedule. The last three lines of the table show a few more of the regular students feeling that these activities took too much time and energy and slightly more of the accelerates feeling that they were too frivolous, both groups agreeing that they were of value.

In considering Table 4 it is important to keep in mind that the accelerates and their paired regulars have been out of school the same length of time but that the accelerates were approximately a year younger, since accelerates and controls were selected as graduating in the same year and entering at the same age but the accelerates had been in school one year less. Thus in the same time since the undergraduate degree and though a year younger, twice as many of the accelerates have earned a further degree. In fact, the accelerates have earned thirty-two degrees: twenty-four Master's, one L.L.B.; one M.D.; four Ph.D.'s; and two Physical and Occupational Therapy certificates. In contrast, the only further degrees the regulars obtained were twelve Master's.

Slightly more of the regulars are married, slightly more of those married have children; slightly more of the accelerates have been divorced or separated. The accelerates married about a year younger; being as a group a year younger, they have been married about the same time. Whether such slight differences in the mar-

TABLE 4.—FURTHER SCHOOLING, MARITAL STATUS, EMPLOYMENT, AND COMMUNITY ACTIVITIES

	Acc.	Reg.
Percentages who have had further schooling	51	43
Earned further degrees	24	12
Now married	80	85
divorced or separated	5	2
median age of marriage	22.3	23.2
percentage of married having one or more children	83	90
single	14	13
Occupational status: housewife	71	83
employed (some also housewives)*	39	28
In at least one community activity (as church)	84	84
In three or more	44	49

* Of the married accelerates, 29% were employed; of the married regulars, 16%.

riage, divorce and child-bearing rate are significant, especially in view of the differences in age, can hardly be judged. Slightly more of the regulars are housewives, and appreciably more of the accelerates are employed—substantially more of the married accelerates are also working. The numbers in community activities are approximately the same. In short, the two big differentiae are the substantially greater number of accelerates obtaining advanced degrees and substantially greater number continuing some sort of a career after marriage. The accelerates included a physician, an attorney, three chemists, a bacteriologist, etc.

As mentioned earlier, there were in various places opportunities for the respondents to comment, and a special invitation so to do at the end of the blank. The comments were various. Some accelerates felt that the program, though desirable, was so only in an emergency. Others were delighted at the saving of time, and considered that such possibilities of time-saving should always be available. Some complained of summer attendance, but others liked the contact then with more mature students. Accelerates who had been out in full-time employment tended to like the businesslikeness of a four-quarter schedule; one remarked that simply the four ten-day periods between quarters gave her more vacation than she had had on the job. Some complained of heavy work, but others liked it, and said they worked better under some pressure.

There were sundry comments about curricula and instructional methods. Several alumnae wished that, somewhere, they had been taught to type; a number of the married women wished they had had courses in homemaking and child care. Some of these young women felt that their curricula should have prepared them more adequately for their vocation. Others wished that their programs had been broader, and, in fact, a number of the regulars had 'accelerated' a bit (for instance, gone to a summer session) to enrich their schedules with extra electives; three had obtained two undergraduate degrees in four years. It was pleasant to find, in this time of widespread criticism of schools and colleges, that many of the comments were favorable; most alumnae looked back upon their college years as pleasant and profitable.

GENERAL SIGNIFICANCE OF THE FINDINGS

The present findings regarding acceleration, as viewed by students ten years after graduation, seem in substantial accord with results obtained earlier, and by other investigators (*1*, *3*, *4*, *5*). It seems clear that many students of good ability can complete the usual four-year undergraduate program in less time without any unfortunate outcomes. Several in the present study so finished in two calendar years. Further, most of the accelerates in this study (and most others) chose accelerated programs and carried them forward without any special counseling in the matter; a few should have been warned away from acceleration, and a majority might have planned better. The writers have found that with careful selection and guidance, outcomes are even better, as has the Fund for the Advancement of Education (*1*, *4*). Especially does it seem desirable that more use be made of credit by examination, for secondary school work beyond entrance requirement, also sometimes for crediting independent study (*4*, *7*). But the question now is as to how such findings are to be explained.

The accelerates in the present study were not notably superior in general ability to the average graduating student—they had tested only about ten percentiles higher on the test of general ability given at entrance to college. The possibility comes to mind that perhaps many undergraduate programs are sufficiently easy, studded with vacations, sometimes repetitious that some time-saving could become more general, or the programs be made more substantial.

But the accelerates in important respects actually did better than the controls: in academic record, number obtaining advanced training, and number continuing a career. If they had been men, presumably yet more of them would have thus pushed ahead. Previous research of the writers, and the findings of Terman regarding his gifted group, indicate that accelerates are more likely to be successful and even distinguished in their careers (*4*, *5*). How are these anomalous findings to be explained: better schooling in less time, and better career too?

In most studies above referred to, the accelerates were equated with controls in initial ability, but not in ambition and energy. The person with more drive may more often accelerate, and work harder in school and job. Certain other studies have been broad enough or so planned that they somewhat allow for these factors, and still the accelerates did better. Might acceleration in itself have beneficial effects?

The accelerates not infrequently mentioned the challenge of such an intensive program, and the stimulus to confidence and ambition resulting from success in it. An opportunity into desired advanced training and career a year sooner can be very appealing to able young people. Earlier graduation may obviously leave more time and funds for further training. Certain other possible factors seem worth mention. May it perhaps be that in the growth years the organism is in subtle respects more ready for the broad learnings of undergraduate curricula, that as it becomes mature it is constitutionally set for adult career, and that careers begun early enough to be fully energized by the vigor of young adulthood are most likely to be successful? These last are indeed hypotheses. But there seems much to suggest that they have some soundness (*2*, *4*, *5*, *6*).

SUMMARY

This paper reports a follow-up, some ten years after graduation from college, of a group of young women who during the war years completed a four-year undergraduate program in three years or less, and of cases then paired with them as to year of graduation, ability and age at entrance, and nature of college program. A previous study had shown the accelerates doing slightly better academic work than their controls and participating about as much in campus activities. The question was as to after-graduation careers of the two groups, and their retrospective judgments about their college programs.

A brief inquiry form was mailed to all cases (one hundred and thirty-nine accelerates and one hundred and forty regulars) for whom addresses could be found. Returns were obtained from one hundred and twelve accelerates and one hundred regulars (eighty-one per cent of the first and seventy-one per cent of the second group).

The one-time accelerates reported almost as much employment while in school; and few accelerates or regulars felt that such employment unfortunately affected health, schooling, or social life. Most thought such work experiences of value.

Most common method of acceleration was a four-quarter schedule, but a considerable number of students also obtained credit by examination, took heavy course loads, or used some combination of the above three methods. Very few accelerates thought their accelerated programs unfortunately affected health, school work, or social life. A majority felt acceleration a desirable challenge and saving of time. The non-accelerates were throughout more critical and dubious about acceleration.

About as many accelerates as non-accelerates took part in extracurricular activities related to the war, but somewhat fewer participated in the usual undergraduate activities. Since there are few such in the summer, and most are organized on the presumption of a four-year program, the accelerates might be considered participating relatively well.

Slightly fewer of the accelerates have so far married and had children, but it must be kept in mind that they are a year younger than the regulars. The accelerates married almost a year younger, or about the same time after graduation. About the same number in each group are now active in community affairs. Twice as many of the accelerates have earned further academic degrees, as compared to the non-accelerates. Almost twice as many of the married accelerates are continuing a career.

The above findings are explained as possibly in part due to a greater ambitiousness and energy on the part of the accelerates, but probably also a result of certain challenges and advantages of accelerated programs. The considerations involved seem of fundamental importance in planning secondary and higher education, and more so for men and in furtherance of career than for women as in this study.

REFERENCES

1) *Bridging the Gap Between School and College*. New York: The Fund for the Advancement of Education, 1953.

2) H. C. Lehman. *Age and Achievement*. Princeton: Princeton University Press, 1953.

3) G. Melcher. *An Evaluation of the Northeast Accelerated Junior College*, Kansas City, Mo.: Board of Education, 1953.

4) S. L. Pressey. *Educational Acceleration: Appraisals and Basic Problem*, Columbus, Ohio; Ohio State University Press, 1949.

5) L. M. Terman and M. H. Oden, *The Gifted Child Grows Up*. Stanford, Calif.: Stanford University Press, 1947.

6) L. M. Terman. "The discovery and encouragement of exceptional talent," *American Psychologist*, 1954, 9, 221–230.

7) Mazie E. Wagner. *Anticipatory Examinations for College Credit: Twenty Years Experience*. Buffalo: University of Buffalo Studies, Vol. 20, No. 3, December, 1952.

A NOTE ON THE STANDARD ERROR OF MEASUREMENT

C. H. PATTERSON

Veterans Administration Regional Office

St. Paul, Minnesota

The standard error of measurement of a test score has sometimes been suggested as being more useful in counseling the individual than other measures of reliability. There appears to be some disagreement on the meaning of this statistic, and thus of the nature of the statements which are made concerning it. This problem was aired in the exchange of papers between Guilford and Garrett in 1937 (*1*, *2*, *4*), in which no apparent agreement was reached.

Two main types of difficulty appear in interpreting the standard error of measurement. The first concerns the nature of the statements which can be made about the true score of an individual with a given obtained score, in terms of the standard error of measurement. The second difficulty, more general in nature, concerns the type of statement which can be made in any situation where probabilities are involved.

The first problem basically involves the meaning to be assigned to an obtained score on a test, and its relation to the 'true' score corresponding to a given obtained score. Some writers, e.g., Guilford (*3*) and Lindquist (*9*) state that no exact statement can be made concerning the true score of an individual making a given obtained score. They insist that the standard error of measurement applies to the observed score, and not to the true score. Since the standard error of measurement is the standard deviation of the distribution of the scores which an individual would make on repeated tests, statements must be limited to these obtained scores. The mean of this distribution of obtained scores would be the individual's true score. Thus, a typical statement would be: "If a certain individual has a true score of 50 points in the test, then we may expect two-thirds of all his actual scores to lie between 46 and 54" (*3*, p. 414), given a standard error of measurement of 4. Lindquist (*9*, pp. 221–222) follows a similar interpretation. Gulliksen also agrees that "no probability statements can be made to apply to all persons who make a given *observed* score," but only for all persons with a given true score.

These writers suggest that the only way in which statements about true scores may be made is by means of the standard error of estimate of a true score predicted from an obtained score by means of a regression equation. This method is seldom used because of the labor involved.

Another group of writers make statements concerning the true score on the basis of an obtained score and its standard error (e.g., *7*, *10*). They do this by taking the obtained score as the estimate of the true score, either explicitly or implicitly, and using the standard error of the observed score as being equivalent to the standard error of a true score. Kelley writes: "σ_{1_∞} may correctly be used to indicate the standard error when x_1 (an obtained score) is estimated from a knowledge of x_∞ (a true score) (a process actually never performed) or when x_∞ is estimated by using non-regressed x." (*8*, p. 403). This method leads to interpretations such as the following (given a score of 70 and a standard error of 7.72): "There is one chance in six that John's true ability lies below 62.3, so that there are four chances in six that it lies between 62.3 and 77.7" (*7*, p. 172).

Some writers make statements of the type that the true score of an individual with a given obtained score is 'almost certain' to lie within three standard errors of the obtained score. Peatman, for example, referring to confidence limits of an IQ of 109, given a standard error of 4.5, states: "thus it is likely that persons with an IQ score of 109 have parameter (i.e., true) scores whose values lie between 100.0 and 118.0; and we can be quite confident that the parameter values will not be less than 95.5 or greater than 122.5" (*11*, p. 379). These statements do not seem to be exact probability statements, though they may frequently be taken as such by readers. They thus do not appear to be of the type which implicitly accept an obtained score as an estimate of the true score. They may be of a type described by Gulliksen (*6*, p. 20) in which 'reasonable limits' are assigned to the true score.

While the observed score is an estimate of the true score, it is not the true score, nor the best estimate of the true score. It would thus appear that statements based on the acceptance of the obtained score as equivalent to the true score are questionable. Statements concerning the true scores of individuals with a given observed score should strictly be based on the regressed score and its standard error of estimate. While the effort involved in this procedure of estimating true scores may not be justified, statements concerning the true scores associated with given observed scores should probably be limited to regressed scores. Rigorous logic would indicate, therefore, that statements should be of the type used by Lindquist, Guilford and Gulliksen.

Nevertheless, it appears that the use of the obtained score as evidence of the true score is common. The correct form of statement when this is the case is of the form that for a given observed score the corresponding true score will fall within a given interval, determined by the observed score, its standard error, and the confidence level desired, and we shall be right in this statement in a specified proportion of cases in the long run corresponding to the confidence coefficient selected.

This leads to the nature of probability statements, regarding confidence intervals within which a given parameter, such as a true score, lies. The true score is a fixed value, and not a random variable. Therefore, in a particular case it either is or is not within a given confidence interval, i.e., the probability is either unity or zero, respectively. Probability should be attached to the statement, which is true in a given proportion of the cases in the long run, and not to a score. It cannot be stated, therefore, that the probability is two-thirds, or 68 out of 100, that a given individual's true score lies within plus or minus one standard error of his observed score taken as an estimate of his true score. The correct statement is that a given individual's true score lies within plus and minus one standard error of his observed score, and we shall be correct in this statement for two-thirds of the cases, or approximately 68 times out of 100.

Following the frequency theory of probability, statements must be of this type, the probability applying to the truth or falsity of the statement in a series of cases, not to a score. Following the interpretation of Guilford, Lindquist and Gulliksen, the correct statement would be that a given individual's observed score lies within plus and minus one standard error of his true score, and we shall be correct in this statement in approximately 68 cases out of 100. The correct form of statement regarding a true score estimated by means of the regression formula would be that for all cases with the same estimated true score, the actual true score falls within three standard errors of the estimated score over ninety-nine per cent of the time.

It is possible to make certain statements about the scores of groups of individuals. Using the concept of tolerance limits, it appears to be possible to assign exact probabilities to statements referring to a group of individuals with a given obtained score.[1] Usually, however, it is sufficient to be able to state that in repeated testing the obtained scores of a group of individuals with a given obtained score on the first test will almost certainly fall within three standard errors ('reasonable limits') of the corresponding true score. It is possible to state that "for all persons with a given true score, the probability is over .997 that the observed score will lie between plus or minus three times the standard error of measurement for that true score" (*6*, p. 20).

While there may seem to be little practical significance to the differences in statements discussed above, it is important that those using the standard error of measurement understand the nature and the bases of the different types of statements. It is hoped that this brief discussion will serve to clarify these differences, which, to the knowledge of the writer, have not been clearly differentiated heretofore.

REFERENCES

1) H. E. Garrett, "On the interpretation of the standard error of measurement." *Amer. J. Psychol.*, 1937, 49, 679–680.

2) — — — "A rejoinder." *Amer. J. Psychol.*, 1937, 49, 683–685.

3) J. P. Guilford. *Psychometric Methods*. New York: McGraw-Hill, 1936.

4) — — — "More concerning the interpretation of the standard error of measurement." *Amer. J. Psychol.*, 1937, 49, 680–683.

5) — — — *Fundamental Statistics in Psychology and Education*. (2nd Ed.) New York: McGraw-Hill, 1950.

6) H. Gulliksen. *Theory of Mental Tests*. New York: Wiley, 1950.

7) T. L. Kelley. *Interpretation of Educational Measurements*. Yonkers, N. Y.: World Book Co., 1927.

8) — — —. *Fundamentals of Statistics*. Cambridge: Harvard University Press, 1947.

9) E. F. Lindquist. *A First Course in Statistics*. (Rev. Ed.) Boston: Houghton Mifflin, 1942.

10) Q. McNemar. *Psychological Statistics*. New York: Wiley, 1949.

11) J. G. Peatman. *Descriptive and Sampling Statistics*. New York: Harper, 1947.

---

[1] For the method of doing this, see "The interpretation of the standard error of measurement," by the author in the March, 1955, issue of the *Journal of Experimental Education*.

THE STABILITY OF INTEREST SCORES OF COLLEGE FRESHMEN

CLAY E. GEORGE and ALBERT J. KINGSTON

The Agricultural and Mechanical College of Texas

It is a common practice among colleges and universities to administer some type of psychological tests to entering freshmen. Many of these instruments are designed to evaluate a student's aptitude for academic work or to appraise his achievement in certain subject matter areas. The results of these tests often are employed for placement purposes. Some institutions also administer an interest inventory at the same time, the results of which often are utilized during subsequent counseling throughout the student's freshman year. The research of Strong (*3*) has indicated that the interest patterns of young men of college age tend to fluctuate considerably. As it is axiomatic in counseling that a counselor must have confidence in any test results which he employs, it was felt that a study of the stability of the interest scores of college men might be valuable.

Three approaches were decided upon. First, the test re-test means were compared for each of the ten scores of the *Occupational Interest Inventory* (O. I. I.) (*1*). Next, the profiles (pattern) of the scores were tested by computing rho's between the test re-test scores for a large number of students. The final approach was to make a frequency distribution of the number of students scoring higher or lower on the nine fields and types of interest to various degrees.

METHODOLOGY

On September 5, 1953, all entering freshmen at the Agricultural and Mechanical College of Texas were administered the *Occupational Interest Inventory*. On October 27, 1953, a randomly selected group of two hundred fifty-six male students were readministered the inventory.

The major purpose of the study was to determine the stability of individual patterns rather than group patterns. In order to obtain this information three different techniques of analysis were employed. First, the means and standard deviations for each field

TABLE I.—TEST RE-TEST MEANS AND STANDARD DEVIATIONS FOR FIELDS, TYPES, AND LEVELS OF INTEREST

Scale	Pers. Social	Fields of Interest					Types of Interest			
		Nat.	Mech.	Bus.	Arts	Sci.	Verbal	Manip.	Comp.	Level of Interest
Test Mean	14.86	22.10	22.46	19.00	13.45	25.77	7.62	12.45	9.81	69.95
SD	5.12	6.50	6.29	8.15	9.54	6.41	4.44	2.49	3.95	10.12
Re-Test Mean	14.96	22.33	22.19	20.63	15.99	24.59	7.73	11.75	9.45	69.76
SD	6.68	9.79	5.48	8.14	7.74	7.52	4.73	3.71	4.31	8.75
D	.10	.23	−.27	1.63	2.54*	−1.18	.11	−.70*	−.36	−.19

* P = .05

and type of interest, as well as the level of interest, were determined and compared. Table I illustrates the relationships obtained by this comparison.

This analysis seemed to indicate that there was little difference between the average scores of the group on the initial test and on the re-test. Only two scores seemed to be significantly different—the Arts score, showing a slight gain, and the Manipulation score, showing a slight loss. Both were significant at the 5 per cent level of confidence.

In order to determine whether there were significant changes among the various factors which comprised a student's pattern of interests, the scores of one hundred selected men were analyzed. Rho coefficients were calcuated between the percentile scores of each field of interest, type of interest, and level of interest obtained

TABLE II.—RANK-DIFFERENCE CORRELATIONS BETWEEN PERCENTILE SCORES ON THE TEN INTEREST SCALES OF THE O.I.I. FOR 100 RANDOMLY SELECTED MALE FRESHMEN

No. of Students	Rho	P
3	−.20– .00	N.S.
17	.01– .54	N.S.
11	.55– .64	<.10
14	.65– .74	<.05
55	.75–1.00	<.01

TABLE III.—CHANGES IN PERCENTILE RANKINGS OBTAINED BETWEEN FIRST AND SECOND TEST (FIELDS AND TYPES OF INTEREST)

	No Change in Score	Percentile Score Increase					Percentile Score Decrease				
		1–19	20–39	40–59	60–79	80+	1–19	20–39	40–59	60–79	80+
Pers.-Soc.	68	32	30	7	1	2	49	57	8	1	1
Nat.	87	58	29	8	0	0	52	19	1	2	0
Mech.	73	55	34	4	0	0	39	46	5	0	0
Bus.	76	51	34	7	1	0	41	39	6	1	0
Arts	63	57	47	10	0	0	49	26	4	0	0
Sci.	66	46	24	4	0	0	47	44	20	4	1
Totals	433	299	198	40	2	2	277	231	44	8	2
Verb.	72	36	44	8	3	0	44	36	13	0	0
Manip.	50	40	53	13	2	0	42	41	13	1	1
Comp.	70	38	29	6	1	1	46	52	11	2	0
Total	192	114	126	27	6	1	132	129	37	3	1

on the initial test and the re-test. The rho's for each of these one hundred students are given in Table II. The results of this analysis indicated that the general pattern of interest scores had undergone considerable change in at least twenty cases.

This analysis would seem to indicate that the interest patterns of at least twenty of the one hundred students had fluctuated so widely that use of the initial test would prove to be very misleading.

One further technique of analyzing the data was employed. It was felt that analysis of the changes in each interest scale according to the percentile rankings provided by the authors of the inventory would prove helpful in illustrating the degree of fluctuation found in some individuals. Table III illustrates these findings.

CONCLUSION

This study may be regarded as substantiating the axiom that test results must be interpreted with caution. The measured in-

PRELIMINARY EXPLORATION OF PERSONALITY FACTORS RELATED TO SUCCESS IN A HIGH-LEVEL MILITARY SCHOOL

JAMES H. L. ROACH and EDWARD M. WALL

The Air University

In the Air Command and Staff School, as in other educational institutions, we have found that intellect and achievement are far from perfectly correlated. One of our best predictors of success in this course ($r = .64$) has been the Initial General Survey Test (IGST) which was constructed to measure military and general background knowledge.

However, in the AC&SS class '53-B there were fourteen officers who fell in the 'C' group on the IGST but whose achievement in the course was out of line with expectancy. Of these fourteen, six ended in the 'A' group academically (top seven per cent) and eight in the 'E' group (bottom seven per cent).[1] When it was further noted that the ACE (American Council on Education Psychological Examination) scores of these two groups were essentially the same, their differences in achievement became even more interesting. An attempt was then made to find if there were any measurable personality differences between the two groups which could account for the differential.

Faculty advisors of these students, who had worked with them closely, were asked to rate their students on a specially devised scale listing twenty variables. Scores on each item ranged from zero to five. Advisors were not told why these students were selected for this rating.

Moreover, during the course, fellow students had filled out officer effectiveness report forms on each other as a part of their regular course program.[2] From these reports about twenty-five independent terest patterns of certain male freshmen had changed radically after a period of less than two months of college work. It seems likely that a counselor who relied on the results obtained at the initial inventory would err in interpreting the interests of many counselees.

---

[1] Final standing in the course was based upon four tests and a final examination, two seminar grades, writing and speaking grades and a staff study report. Large, significant differences between the two groups of subjects were found on all of the above factors except the staff study report.

[2] Student officers were required to complete these forms on each fell[illegible]

This study also illustrates the error one may make by relying solely upon test re-test averages in studying score stability. From the averages given in Table I, it would seem that there was very little change in either pattern shape or subtest levels. Tables II and III indicate, however, that there were a very large number of marked individual changes in both profile shape and level.

Reid (*2*) conducted a similar study with the *Kuder Preference Record*. He reports rank order coefficients ranging from −.72 to 1.00, with a median of .88. These were based on freshman college student interest patterns as measured at entrance and fifteen months later. In the present study the range of rhos was −.20 to 1.00 with a median of .77. Strong (*4*) reports a correlation of .81 between permanence of interest scores and time interval between test and re-test, with the influence of age partialed out. Since the Kuder tends to be more stable over a fifteen-month period than the Lee-Thorpe over a two-and one-half-month period, and permanence decreases with time, it would appear that the Kuder is the more stable instrument. Sampling differences may have affected these results, however, so this conclusion is only tentative.

REFERENCES

1) F. A. Lee, and L. P. Thorpe. *Occupational Interest Inventory*. Advanced Series. Los Angeles: California Test Bureau, 1943.

2) J. W. Reid. "Stability of measured Kuder interests in young adults." *J. educ. Res.*, 1951, 45, 307–312.

3) E. K. Strong. *Vocational Interests of Men and Women*. Stanford: Stanford University Press, 1943.

4) — — — "Permanence of interest scores over twenty-two years." *J. appl. Psychol.*, 1951, 35, 89–91.

TABLE I.—MEAN SCORES OF OVER-ACHIEVERS AND UNDER-ACHIEVERS ON PERSONALITY FACTORS WHERE SIGNIFICANT DIFFERENCES WERE FOUND IN FACULTY RATINGS

Factor	Mean Score Over-achievers	Mean Score Under-achievers	Differences in means
Originality	4.20	2.00	2.20
Initiative	4.18	2.15	2.03
Judgment	4.18	2.23	1.[illegible]5
Personality	4.18	2.31	1.87
Personal adjustment	4.22	2.38	1.84
Activity level	4.27	2.62	1.65
Emotional stability	3.73	2.23	1.50
Leadership	3.64	2.23	1.41
Reliability and Responsibility	4.55	3.15	1.40
Sports activities	4.00	3.08	.92

TABLE II.—MEAN SCORES OF OVER-ACHIEVERS AND UNDER-ACHIEVERS ON PERSONALITY FACTORS WITH SIGNIFICANT DIFFERENCES FOUND IN RATINGS OF FELLOW STUDENTS

Factor	Mean Score Over-achievers	Mean Score Under-achievers	Differences in means
Leadership	11.17	8.00	3.17
Growth potential	11.17	8.38	2.79
Reliability and Responsibility	11.33	8.75	2.58
Over-all evaluation	8.67	6.13	2.54
Judgment	11.33	8.88	2.45
Coöperation	11.33	9.75	1.58

willingness to work in harmony with others), and growth potential (capacity to handle jobs of increased scope and responsibility, self-improvement efforts, ability to learn rapidly). The one variable rated by students and not faculty was over-all effectiveness, e.g., a comparison with all officers in the same grade.

Results of this study show clearly that certain favorable personality characteristics were exhibited to a much greater degree by the over-achievers than by the under-achievers. Inasmuch as these two groups were about equal in intellectual capacity as well as background knowledge, it is felt that these personality differences, especially leadership, judgment, reliability and personal responsibility, account in part, at least, for the superior success of the over-achievers. These findings are in general agreement with those re-evaluations on six variables, (five of which appeared on the faculty's scale), were available for each of the fourteen subjects. Scores on this scale ranged from one to sixteen. (A five-point scale each broken into three subdivisions with one subdivision at the bottom end of the scale.)

The faculty rated the 'over-achievers' higher on each of the twenty factors; and for ten of these the differences were statistically significant (Table 1). The ten significant factors with general description were as follows:

1) Personality: closed, withdrawn or open, pleasant.

2) Emotional Stability: tense, upset, involved, or free from 'blowups'.

3) Leadership: ability to command, exert control, inspire confidence; skill in directing others.

4) Activity level: degree of participation in general activities.

5) Reliability and responsibility: degree to which accepts responsibility and can be relied upon to execute assigned tasks.

6) Personal adjustment: problems faced and ease in solving them.

7) Judgment: ability to grasp a situation, soundness and logic of thought.

8) Initiative: frequency of initiating and explaining projects and ideas.

9) Originality: creativeness and originality of thinking.

10) Sports activities: degree and quality of participation.

Factors in which the differences were not statistically significant were: persistence, general health and physical well-being, family adjustment, job knowledge, coöperation, growth potential, motivation, tact, study habits, and social activities. Oddly, faculty members reported that they had more difficulty in remembering the better students.

Evaluations by fellow students also gave the over-achievers higher ratings on all six of their variables, but here all differences were statistically significant (Table II). Student ratings on judgment, leadership, and reliability and personal responsibility agreed closely with faculty ratings. However, students rated the over-achievers higher than the faculty did on coöperation, (ability and

---

officer in their seminar to: (1) Learn how to complete the form properly; and (2) acquaint students with an appraisal of their effectiveness as seen by fellow students. They were not used in course grades.

ported by Terman, Cox, and others in respect to personality factors related to life success of gifted individuals.

REFERENCES

1) Catharine C. Cox. *The Early Mental Traits of Three Hundred Geniuses.* Vol. II of Genetic Studies of Genius, L. M. Terman, (Ed). Stanford: Stanford University Press, 1926.

2) L. M. Terman, and M. H. Oden. *The Gifted Child Grows Up.* Vol. IV of Genetic Studies of Genius. L. M. Terman, (Ed). Stanford: Stanford University Press, 1947.

BOOK REVIEWS

STEPHEN M. COREY. *Action Research to Improve School Practices.* New York: Bureau of Publications, Teachers College, Columbia University, 1953, pp. 161. $3.00.

Dr. Corey's proposal for action research springs from his conviction that dramatic changes must be made in public school curricula and practice if youth are to be prepared for the demands of a changing world. He believes that the knowledge gained from research is most likely to be translated into action if teachers and school administrators are ego-involved in the research.

The quantity of space given to the value and processes of group decision may surprise the reader, but the author emphasizes the necessity for enlisting the coöperation of all who will be affected by the changes. Reports of some actual projects are given to illustrate the steps in formulating and testing hypotheses.

Some distinguishing characteristics of action research are stated: it is focused on practical problems facing the participants; it is carried on amid the complexities of the school without the customary controls of the laboratory experiment; its value is measured in terms of improvement in educational practice; research design is subject to modification during experimentation; generalizations from research are extended vertically instead of laterally. It goes further than casual inquiry because evidence is systematically recorded, analyzed and interpreted.

Participants, Corey says, must be willing to admit and explore flaws in the school system, recognize that not all hypotheses will be proven, develop skills in leadership and group membership, evaluate group meetings, try to understand the feelings of colleagues, and provide time, resources and adequate consultant services. The last-named essential condition for success, the reviewers feel, may be the most difficult to arrange.

Chapter 7 presents some statistical evidence to support the vertical extension of generalizations and describes some basic statistical operations which might be used in action research. The author believes that the teacher will learn to use these techniques when he becomes aware that they will help him solve problems and answer questions.

The book is addressed principally to school administrators and supervisors whose leadership and support would significantly influence the success of action research. Those who want to test practices and attack persistent school problems will find some useful information. Some readers will probably comment that more attention is given to improving staff harmony than to planning research on instructional problems. Others will infer that the working relations of faculty members will be reflected in the classroom and that principles and techniques presented in this connection can be applied to investigations of the school offering. Some published studies of action research and group decision are summarized by the author, and eighty references are included in the bibliography.

MAURICE D. AND JEANNE A. WOOLF

Kansas State College, Manhattan

SIR CYRIL BURT. *The Causes and Treatment of Backwardness.* New York: Philosophical Library, 1953. pp. 128.

American students of the problem of mental deficiency are familiar with Dr. Burt's contribution in this area. Particularly, are they familiar with the book published fifteen years ago called *The Backward Child*, where the problems of contributions of heredity and environment were considered more sanely and comprehensively than some of the American volumes of that period. Burt considers the present smaller volume an abridgment of this larger one. In abbreviating, Burt has humor enough to quote Boswell that shortening a book was too often "like cutting off the head and the tail of a cow," to which Dr. Johnson was supposed to have replied, "At its best it should be like making the cow produce a calf." Burt hopes that the book is not a simple curtailment, but that it is in many ways a fresh book with a different purpose. The purpose is to serve the staff of the National Children's Home with knowledge and information they can use in dealing with something like 3,000 boys and girls, housed in forty branches.

The suggestion for the book grew out of the lecture that Burt made at one of the Annual Convocation Lectures. This he was asked to enlarge and rewrite in book form. Material of the book is presented in eight chapters dealing with historical factors, methods of investigation, environmental factors, personality factors—physical, intellectual and emotional—and some practical conclusions. The book contains a list of references and a separate index of authors and subjects.

The references used and all other points of reference are distinctly English-bound. This is understandable enough, since the setting and the history is also English. The aim of the book obviously is to search in psychology for knowledge that can be of help in dealing with backward children. One thing that Burt likes to make clear is that the problem is not one of 'physiology of the brain'. The ordinary man, says Burt, erroneously supposes that the expert on physiology is the physician—hence arises the widespread conclusion shared by a number of doctors that psychology is or ought to be a branch of medicine. The problem is not merely a dispute about words; the notion that educational subnormality is the result of a physical defect, which can be cured by glandular extracts, brain surgery, or a tonic for the nerves is widespread. Then Burt argues that the proper field, of course, is psychology, and that the first investigations were distinctly English. The investigator who first introduced the scientific approach to the psychology of the individual child was Sir Francis Galton. And it was Galton, Burt insists, who advocated that all such conclusions needed to be checked by new experimental and statistical techniques. To this end he invented an experimental device of mental testing and the statistical method known as correlation. The idea of mental testing, even in a child guidance clinic, according to Burt, was borrowed by Binet from Galton, and was introduced in America by Cattell, who was one of the first of Galton's fellow-workers. All of which is to prove that the kind of psychology that can help backward children originated in England and its most profound influences came out of England.

Nobody in the United States is going to argue against the importance of Galton or the fact that Cattell worked with him. But few, if any, would agree with the particular over-Anglicized emphasis by Burt about the origin of all things good in England that has to do with the study of backward children.

This bias does not detract from the sane and sensible, intelligent approach in considering the problems that Burt makes in this book as he has in his previous publications. For a straightforward presentation of the causes of, and methods of dealing with, backwardness, this book can be of help to the teacher of backward children and students of the field. Stressed by Burt is what one would expect him to stress; namely, the need for records that should be objective so the analysis can be carried out according to a systematic plan, because, as he says in the last sentence in the book: "In this way, and this way alone, will the accumulated wisdom gained by each one be eventually placed at the disposal of all."

H. MELTZER

Psychological Service Center, St. Louis

BRUNO BETTELHEIM. *Symbolic Wounds, Puberty Rites and the Envious Male.* Glencoe, Ill.: The Free Press, 1954, pp. 286. $4.75.

Dr. Bruno Bettelheim, best known for his book about children, *Love Is Not Enough*, has made certain observations on the children in the Sonia Shankman Orthogenic School, University of Chicago, which prompted him to question certain psychoanalytic anthropological theories about circumcision and rituals. Out of these observations on neurotic and psychotic children has come this volume of what may be considered a re-interpretation of puberty rites and circumcision.

This, on Bettelheim's part, represents an attempt to separate fantasies of patients from historical reality. Contrary to the ordinary belief that initiation rites are imposed by society on the young as a means of introducing them into adult society, Bettelheim has come to almost the opposite conclusion; namely, that they are the efforts of children to help themselves and they are not prompted by elders or irrational drives. Instead, they are based on the urge for freedom from those drives and pressures to master the irrational forces. That's the thesis.

The evidence is based on Bettelheim's own observations and a second and a third look at some anthropological reports. These convinced him that there has been an over-emphasis on the assumed end and under-emphasis on the means of obtaining it; that there has been an over-emphasis on penis-envy in girls and castration anxiety in boys and the neglect of what might be considered the vagina-envy in boys—the fascination of boys with female breasts, lactation, pregnancy and child-bearing. This thesis is presented in a series of chapters which, to some extent, represent arguing with himself about observations, theoretical considerations and interpretation of meanings of such material as Monotheism and the Castrating Father; Fertility, the Basic Cult; Castration, Circumcision and the Great Mother; Australian Rites; The Men-Women; The Secret of Men; and Girls' Rites. The author ends the book with an expressed wish that if we could satisfy through rituals, or through more civilized—less magic and more satisfying—institutions, the desires of both men and women to participate in the activities and enjoyments normally belonging to the other sex, each sex would gain greater inner autonomy, could better accept its own rôle and that of the other, that two could live with one another more satisfactorily. Instead of trying to solve these problems through ritual, he suggests we could be equally serious in searching for solutions that are more rational, more effective socially and more satisfying privately.

In some instances reported by Bettelheim, both from his own observations in school, based on verbal reports of children, and some observations on, say, the Australian aborigines, in whom he reports he found an absence of castration anxiety repressions, his emphasis is more justified than the over-emphasis given by psychoanalysts and anthropologists who have accepted psychoanalytic interpretations. Whether a whole book is needed to prove that the opposite theory is correct is a question. There is enough in here for easy reading, but to some extent the book is a stretched-out article. The references given are mostly from the anthropological field. Bettelheim is modest enough to recognize he is out of his field here.

H. MELTZER

Psychological Research Center, St. Louis.

MARVIN D. GLOCK. *The Improvement of College Reading.* Boston: Houghton-Mifflin Company, 1954, pp. 307. $2.50.

The material presented in this text is derived from a college reading program developed and successfully used for several years at Cornell University. It is designed for use in any college reading program and in English courses which emphasize improvement of reading. Stress is placed upon improving comprehension, analytical skill, judgment, skimming ability, vocabulary study, and versatility in adjusting mode of reading to purpose and nature of the material. The importance of motivation is recognized throughout the course.

There are thirty exercises or lessons. In each exercise there are four parts such as skimming, reading to answer a problem, critical reading, and vocabulary development (Exercise 18). There are sets of questions on each part of each exercise. These questions are concerned with facts, understanding of ideas, tone and intent, organizational pattern, and vocabulary. Although speed is measured, comprehension is properly considered to be more important. A particularly good procedure is to check that reading is not too fast for comprehension and to assure that speed is adjusted to purpose and content. Diagnostic charts are provided for recording day-to-day improvement so that the student may check his own gains. Recognizing that a person cannot expect much improvement in vocabulary during a short-term program, emphasis is placed upon vocabulary exercises to acquaint the student with methods of vocabulary study such as the use of basic roots and affixes.

The discussions and procedures in this text are based upon sound experimental findings. The author has been successful in selecting and giving proper emphasis to the basic factors and skills necessary for developing proficient reading. This is one of the best programs of reading instruction at the college level encountered by the reviewer.

MILES A. TINKER

University of Minnesota

THE JOURNAL OF

EDUCATIONAL PSYCHOLOGY

Volume 46 May, 1955 Number 5

A DRAWING TECHNIQUE FOR STUDYING CERTAIN OUTCOMES OF TEACHER EDUCATION[1]

WILLIAM RABINOWITZ

Bank Street College of Education, New York City

and

ROBERT M. W. TRAVERS

Personnel Research Laboratory, Air Force Personnel and Training Research Center, San Antonio, Texas

This article deals with a drawing technique and its success in differentiating students exposed to different teacher education experiences. Evidence of the sensitivity of the technique to changes that take place in the course of teacher education will also be presented.

Students at two teacher training institutions at different stages in their professional preparation were given a blank sheet of paper at the top of which were the following directions: "In the space below, draw a picture of a teacher with a class. Draw as complete a picture as you can. Avoid the use of stick figures. Don't worry about your artistic ability or lack of it; just draw as well as you can."

This article is based upon an analysis of the drawings thus obtained.

---

[1] The research described in the article was undertaken and completed while the authors were members of the Staff of the Office of Research and Evaluation, Division of Teacher Education, Board of Higher Education, New York City.

THE MEANING OF DRAWINGS

There is a growing literature on interpretations of drawings made by clinicians to delineate psychodynamic trends in the cases they study. Drawings have also been used to estimate intelligence in children, especially those with language difficulties. The relevance of these methods of analyzing drawings is marginal in the present investigation. Though drawings have been used as a tool in individual diagnosis our purposes are somewhat different.

If a person is asked to draw a teacher in a class, or to draw a policeman directing traffic, or to draw a foreman supervising a road construction gang, he does not produce a drawing out of a vacuum. The drawings inevitably display some of the ideas he has acquired concerning what these individuals do in these particular situations. This knowledge may be objectively correct or it may be distorted, but that is of no immediate consequence. In a picture of this kind, individuals project their impressions concerning how they think a person might act in the situation specified. Without such projection there could be no drawing. We are not saying that the person who produces a drawing does not exercise some conscious selection over the ideas he presents. In most cases, there undoubtedly is some selection of ideas. However, we have found that the individuals who perform this simple test do not usually appreciate the fact that they reveal their own highly personal ideas about the persons and situations presented.

An aspect of projection which is particularly striking is the fact that in his drawing a student often portrays a teacher who actually looks like himself. A tall teacher is drawn most often by a tall student. Heavy students are likely to draw heavy teachers. The age and sex of the drawer are almost always reflected directly in the drawings. These evidences of projection are most discernible when physical similarities are the object of study. However, projection is not limited to the body image, but extends to a great host of attitudes and conceptions which the student holds.

Many of those with whom the drawing technique was first discussed questioned the procedure since individuals differ greatly in drawing skill. These differences in skill, it was argued, would mean that gifted individuals would have an advantage over those less endowed. Drawing skill would make it possible to represent a teacher in the classroom more completely and attractively. The partial truth of this objection must be admitted. Drawing skill does apparently give an individual greater control over his production. Yet our experience has been that the control manifests itself most often in ways which are irrelevant from the viewpoint of interpretation. It is unimportant whether an idea is well presented or poorly presented. Drawing interpretation focuses attention upon the content of the ideas, rather than its formal representation. The skilled individual has an advantage in those aspects of drawing performance which are of least importance.

TWO DRAWING EXAMPLES

Figure 1 and Figure 2 show two radically different drawings produced by student teachers from the same institution. Both students were young women, both were planning to teach at the elementary school level, and both had taken the same number of credits in professional courses in education. Yet the differences in the drawings are striking. The one represents an informal program in which different children participate in different activities at the same time, and in which the teacher appears to exercise little control over the movements of the children. The other picture represents a more formally structured classroom situation. It seems that the teacher controls the entire situation, all the children perform the same activity at the same time, and the emphasis is on academic learning.

We do not say at this point that these two student-teachers would behave differently when faced with a class. We say only that the drawings portray classroom situations which are radically different. The extent and nature of the differences in the students who produce them is another matter.

Certain additional differences may be noted in these two drawings. In the one drawing, the teacher is the center of the entire picture as well as being the center of the entire learning procedure. In the other one, the teacher is drawn more to the side of the picture and is attending to only a few of the children. In the former one, the teacher is drawn in greater detail and with more definiteness than the children. In the other drawing, no special emphasis is given to the teacher who is drawn in about as much detail as the children. In the one drawing, the only piece of equipment is the blackboard with material on it representing a formal, academic approach to reading. In the other, there is a wealth of equipme

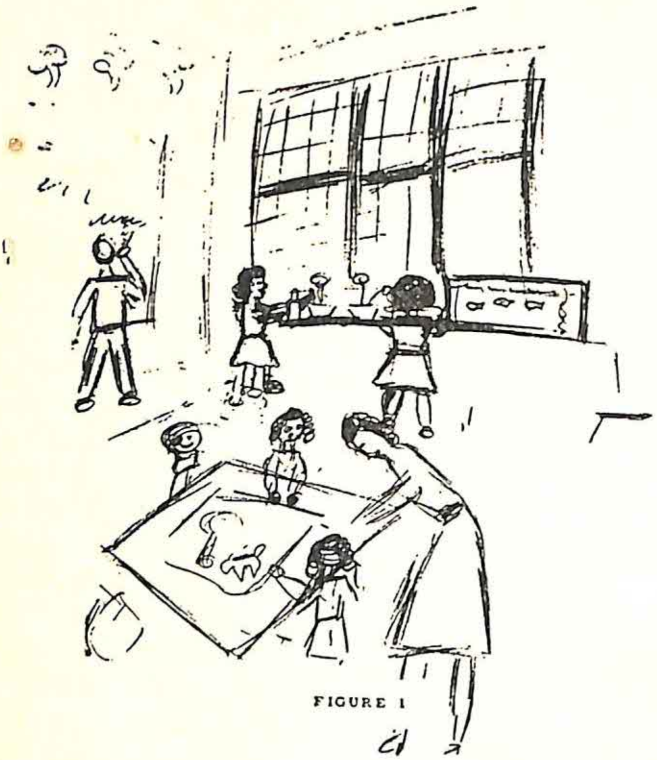


FIGURE 1

and the blackboard is used by the children. The extent to which the teacher is represented as physically close to the children differs in the two pictures. In one there is a gulf between teacher and class, but in the other they are close together.

Certain drawings are extremely bizarre, and even to an observer who has no clinical training they seem 'queer'. Most of them portray some activity, but not an integrated meaningful response pattern in the classroom. The sex of the teacher figure may be impossible to identify. Gross distortions in perspective or the size of various body parts may be present. The drawing is usually permeated by confusion. Drawings of this kind were found in only a very small minority of those examined. In three cases it was possible to obtain informal evaluations of the persons involved given by faculty members over several years. In two cases, the impression of personality disturbance gained from the drawing was directly confirmed by the faculty records. The third case was ambiguous.

In a preliminary examination of the drawings, several students were selected whose drawings before and after student-teaching

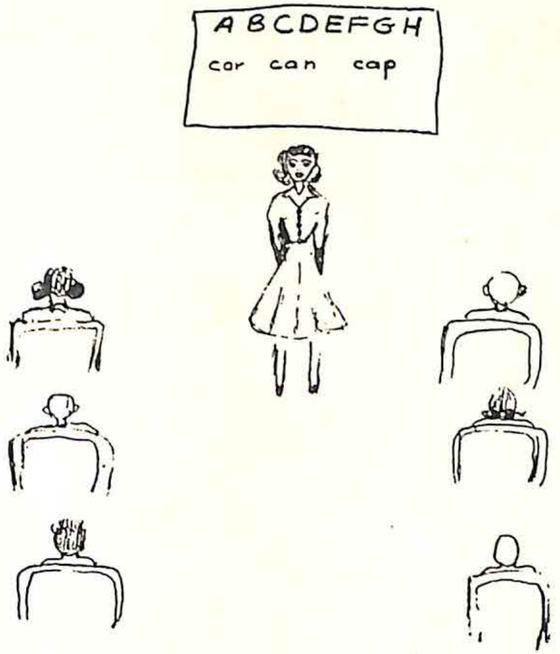


FIGURE 2

showed great changes. Several of these students were interviewed in order to learn if the drawing changes were related to the student-teaching experience. In almost all cases they were.

A QUANTITATIVE APPLICATION OF THE TECHNIQUE

The first part of this paper described a technique for conducting enquiries into the concepts of teaching possessed by an individual. In order to explore further the usefulness of this technique and to investigate the possibility of deriving certain measures from it, the drawings of students of education in two different teacher-education programs were studied. These two programs in Institution A and Institution B differ in many basic respects as is evident from the following descriptions.

DESCRIPTION OF TWO INSTITUTIONS

Institution A.—The program of teacher education in this institution is generally described as being conservative in character. The emphasis on the liberal arts rather than on professional education is reflected in the fact that it does not offer a major in education. The student's major field of specialization must be one of the liberal arts areas of study. Academic scholarship is emphasized. The liberal arts faculty has continually endorsed the viewpoint that a thorough liberal arts training is essential for effective teaching. The program does not include planned observation in selected schools early in the student's training, although students are encouraged to visit classes in an on-campus school. The opinion is commonly expressed that observation is of limited value, or even may be misleading, unless the student has acquired the professional background for interpreting the observations. Thus, a majority of the teacher educational students do not visit public schools as student observers until they first embark on student-teaching. However, this institution has attempted to develop a program of community experiences for its students so that they will have a first-hand knowledge of the community agencies which serve children.

In this institution great emphasis is given to the intellectualistic development of the teacher education student. The program is based upon the assumption that educationally sound behavior of the teacher in the classroom must be based on sound and thoroughly learned professional knowledge.

In this program relatively little is done or can be done to weed out those whose personality seems unsuitable for teaching. Although the speech department does all that can be done to eliminate those who display undesirable behavior during a speech test, the observation of the student in a formal academic setting does not permit the careful and detailed study of individual peculiarities.

Institution B.—In many respects, the teacher education program in Institution B is based on different assumptions from that in Institution A. Professional training is given emphasis in the total program of the teacher education student. It is emphasized that the teacher education student is being trained primarily as a teacher although he is expected to be competent in his subject matter field.

A major assumption underlying the program is that knowledge alone does not guarantee that the teacher will manifest a desirable pattern of behavior in the classroom. Thus, considerable effort is made to provide the student with opportunities to observe in selected schools from the earliest weeks of his professional training. Carefully planned field and laboratory experiences form an integral part of the professional training of the student, and student-teaching represents a final step in familiarizing the student with the problems of educating children.

The schools which the students visit in their programs are not just any schools, but schools which present an educational outlook which Institution B accepts. The development of this program of school visitation and the establishment of relations with suitable schools represents many years of work. Thus, the teacher education students in this program are confronted with a particular pattern of teacher behavior throughout their education.

There is a close relationship between the students and faculty in this institution and this relationship makes it possible for the faculty to study carefully the personalities of the students. On this basis more students are eliminated from the teacher education program. The students themselves seem to be very much aware of the fact that they are being scrutinized and develop anxieties related to whether they will be retained in the program.

These brief descriptions of the two programs should not be taken as an endorsement of the one or the other. Undoubtedly each has its own merits and represents the fruit of serious thought and study of faculties which differ in their values and in their theories concerning how their values can be best achieved. We are not concerned here with problems of value systems but with the relationship of training conditions to concepts of teaching.

PROCEDURE

The basic data in the present study consist of drawings of 'a teacher in a class' made by students in these two institutions who were at various stages of training in the teacher education program. It might have been desirable to follow the same group of students through for a period of two or three years and to have studied changes associated with teacher-training, but the time was not available for undertaking a longitudinal study of this character. Hence, it was necessary to obtain observations on different groups at a number of different levels of training. In order to simplify the study, it was decided to limit it to those planning to become elementary school teachers.

There is no doubt that it would have been desirable to match groups for such variables as age, sex, intelligence, and type of curriculum experienced in high school. In practice, it was, however, not possible to do this. It seems probable that any differences between the two groups on these variables have not greatly influenced the data. The following groups were included in this study:

Institution A	*Institution B*
17 students in Education I (mainly upper sophomores)	17 students in Education I (mainly lower sophomores)
24 students before student-teaching	20 students in Education III (mainly lower juniors)
24 students (the same ones as above) at the end of student-teaching	20 students just after student-teaching

The actual number of students whose drawings were chosen at each level is less than the number who supplied drawings. These students were chosen through random selection. It would have been desirable to have employed larger numbers, but the work of quantification at this stage proved too great. The limited number of cases does not, however, seriously lessen the impact of the results obtained.

QUANTIFYING SOME OF THE CHARACTERISTICS OF DRAWINGS

The drawings are complex representations of complex phenomena. A given pattern of teacher behavior may be represented in drawings in an endless number of different ways. Yet, within this diversity of representation, certain common ideas on specific phases of the teacher's behavior in the classroom are nearly always experienced. For example, the teacher may be separated as the central and major element in the entire drawing or may be so hidden among other objects that he is hard to find. Most drawings will fall somewhere between these two extremes and the drawings can be rated for their position along this scale.

Our initial efforts were devoted to a study of the kinds of characteristics with respect to which consistent ratings could be obtained. This involved the preparation of rating sheets related to characteristics of drawings in which we were interested and the use of these rating sheets by several raters who were graduate students in psychology. The raters were later brought together in conference in order to discover the reasons for discrepancies between ratings. The rating sheet was then revised in order to eliminate those aspects of the rating procedure on which reasonable consistency among raters could not be achieved. The revised rating sheet was again subjected to this procedure which was continued until a rating sheet was developed which seemed adequate for the present purposes. It may be noted that the raters did not know at any time which group they were rating.

The average ratings for each group are presented as points on a graph in order to show trends during teacher education. The results for the three raters have been plotted separately except for the case of the final graph. Tests of significance were applied (1) in order to evaluate the differences between groups at the beginning and end of training and (2) for the purpose of comparing the students in the two institutions at the end of training.

HOW DOES TEACHER-TRAINING AFFECT THE EXTENT TO WHICH THE DRAWINGS SHOW ACTIVITY OUTSIDE OF THE CONTROL OF THE TEACHER AND ACTIVITY SHOWING PARTIAL OR TOTAL PUPIL CONTROL?

Major differences might be expected between Institution A and Institution B in the extent to which the concepts of teaching of the students graduating from these programs included pupil activity outside of the direct control of the teacher. The drawings themselves demonstrated striking differences in this respect as is evidenced from the graphs shown in Figure 3 and Figure 4.

The graphs present data based upon two somewhat distinct ratings. The first, summarized in Figure 3, involves the percentage of drawings in which any activity outside the immediate control of the teacher is portrayed. For example, if a drawing showed a picture of a class in which a formal lesson was being conducted, but with a few pupils at the side or back of the room engaged in some other activity then that drawing would be classified as showing some activity outside of the teacher's immediate control. The second rating, summarized in Figure 4, involves more than this, and indicates the extent to which the drawings portray situations in which there is partial or total pupil control.

The percentage of drawings for the various groups which shows activity outside of the immediate control of the teacher is shown in Figure 3. Changes associated with teacher training are somewhat greater in Institution B than in Institution A. This is to be expected since the concept of teaching which Institution B attempts to develop emphasizes the importance of allowing pupils to engage in activities without direct and constant teacher supervision.

It seems desirable to apply a test of significance in order to determine the probability that the differences between the ratings

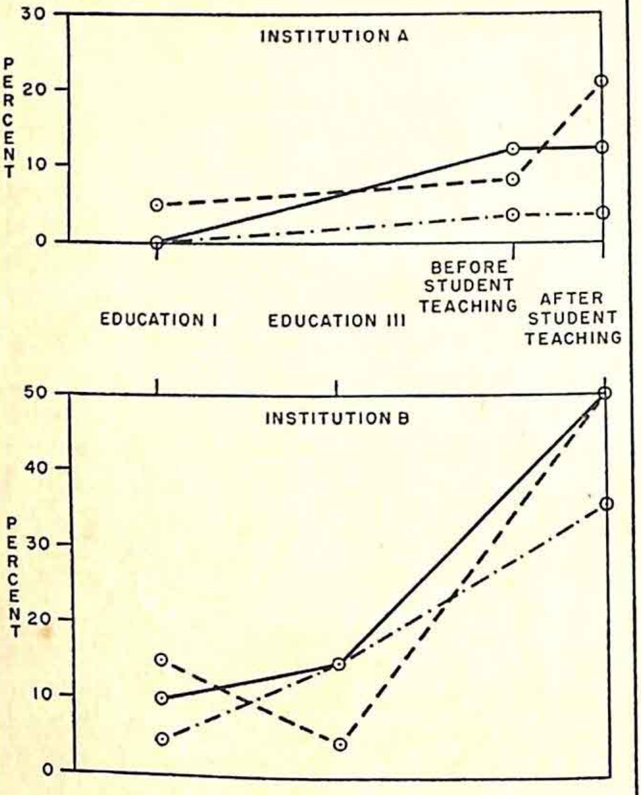


FIGURE 3
PERCENTAGE OF DRAWINGS SHOWING ACTIVITY OUTSIDE OF THE CONTROL OF THE TEACHER

at the beginning and end of training at Institution B are only chance fluctuations. This is done by first drawing up the following four-fold table:

	Number of Drawings Showing Activity Outside of the Control of the Teacher	Number of Drawings Showing No Activity Outside of the Control of the Teacher
Students in Education I	2	15
Students after student-teaching	9	11

It is possible then to apply a chi-square test to determine whether the differences in the way in which the two sets of drawings are

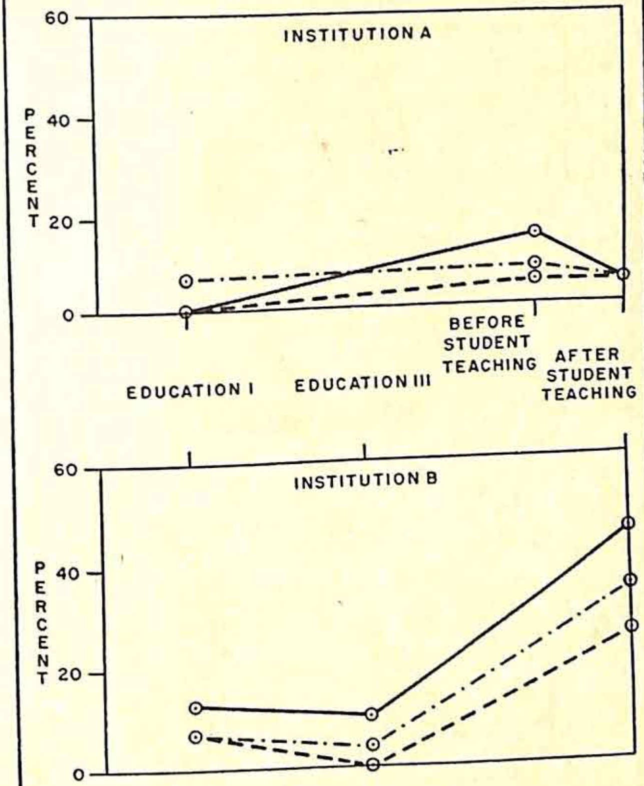


FIGURE 4
PERCENTAGE OF DRAWINGS SHOWING PARTIAL OR TOTAL PUPIL CONTROL

divided should be considered as chance fluctuations.[2] The computed value of chi-square is 4.86 (1 d.f.) which permits us to reject the hypothesis that the differences are a result of chance at the .05 level of confidence.

A similar four-fold table may be drawn up for comparing the drawings made after student-teaching at the two institutions. This table shows a greater difference in the two groups in the fraction of drawings showing activity outside of the control of the teacher. The value of chi-square was found to be 6.94 (1 d.f.) which permits us to reject at the .01 level of confidence the hypothesis that the difference between the samples is due to chance.

---

[2] This four-fold table, and those discussed later, are based on the mean frequencies of the three raters. Yates' correction for continuity was applied in computing all chi-squares.

In Figure 4, the percentage of drawings showing partial or total pupil control of the class activity is presented. Striking differences will be noted between the two teacher-training institutions. In Institution A, no changes are apparent, but in Institution B, a substantial increase is evident.

The difference between the drawings at the beginning and end of training in Institution B may also be tested for significance by exactly the same procedure as was adopted in the case of the previous data. In this case, the value of chi-square is 3.95 (1 d.f.) which permits us to reject at the .05 level of confidence the hypothesis that the difference is one of chance fluctuation. Similarly when the drawings from the two institutions are compared at the end of training the value of chi-square is 6.97 (1 d.f.) which permits us to reject at the .01 level of confidence the hypothesis that the difference is due to chance.

It is significant to point out in connection with the quantitative data presented up to this point that the groups at the beginning of training in both institutions produced very similar drawings, and it has not been possible to demonstrate differences between the beginning groups. The concepts of teaching reflected in the drawings with which students start their training do not seem to differ appreciably from one institution to the other.

HOW DOES TEACHER TRAINING AFFECT THE NATURE OF THE PUPIL-TEACHER RELATIONSHIPS MANIFESTED IN THE DRAWINGS?

The drawings were rated for two major aspects of pupil-teacher relationships. An attempt was made to provide ratings of the extent to which the pictures portrayed tension in the relationship between pupils and teachers. In addition, the raters were asked to identify those drawings in which teacher and pupils were working together on some project.

The average rating for tension in teacher-pupil relations in the two groups is shown in Figure 5. The contrast between these two graphs is striking. At Institution A, there is practically no change in the amount of tension manifested. In Institution B, there is a striking decline in the amount of tension portrayed. The latter change is probably a result of the fact that pupils in the program

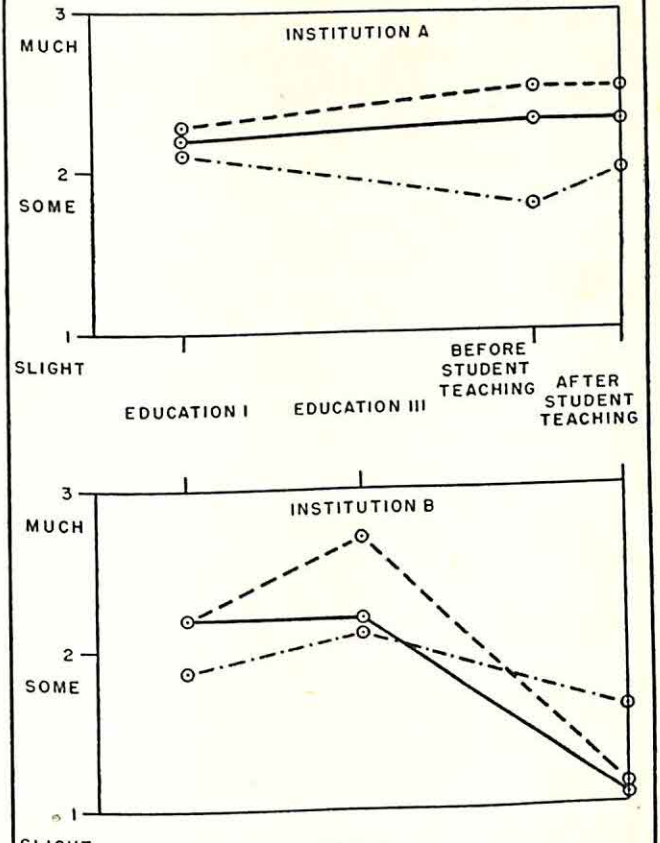


FIGURE 5
AVERAGE RATINGS FOR TENSION IN TEACHER-PUPIL RELATIONSHIPS

have ample and early opportunity to become adjusted to a classroom situation and, hence, to learn adequate responses for handling the situations that arise. At Institution A this process of adjustment can only be begun in student-teaching, and the tensions involved must be resolved in later on-the-job experiences.

In order to test the significance of the difference between the ratings of the drawings at the beginning and end of training at Institution B three ratings for each drawing were averaged. A t test was then applied in order to determine the significance of the difference between the mean ratings for the two groups of drawings. The value of t was found to be 3.546 (d.f. = 35) which p

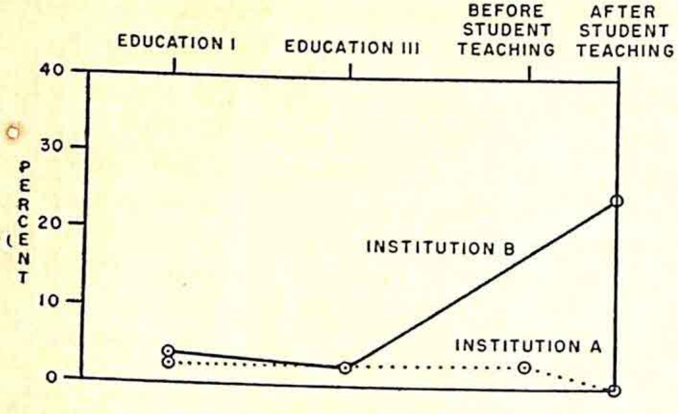


FIGURE 6
PERCENTAGE OF DRAWINGS SHOWING TEACHER AND PUPILS PARTICIPATING IN SOME ACTIVITY

mits us to reject at the .01 level of confidence the hypothesis that the drawings before and after training can be considered as samples of the same population of drawings. Even more certain can one be of the significance between the drawings from the two institutions at the end of training for, by the same technique, *t* is found to be 4.281 for which p = .001.

Figure 6 gives the percentage of drawings which show the teacher and the pupils participating together in some activity. Since the number of drawings which fall into this category is rather small, it seemed desirable to average the percentages for the three raters than to keep them separate. This graph shows the same striking trend as the previous graph. In the one institution, there is an increase in the amount of pupil-teacher participation in common activities; in the other, there is little change and even possibly a decline.

The significance of the difference between the ratings of the drawings made before and after teacher-training at Institution B was determined by the chi-square test in the manner previously described. The value of chi-square for this comparison is 3.82 (1 d.f.) which is significant at the .05 level of confidence. However, one can speak with greater confidence about the difference between the institutions at the end of training with respect to the characteristic rated. In this latter case the value of chi-square is 7.38 (1 d.f.), which is significant at the .01 level of confidence.

VARIABLES HYPOTHESIZED TO BE CONTROLLED BY PROCESSES OTHER THAN TRAINING EXPERIENCES

It is reasonable to assume that some variables are affected by teacher-training experiences. One of them is the extent to which the prospective teacher manifests any psychological disturbance. Presumably, this variable is a function of deep-seated personality characteristics which are not greatly altered in the course of teacher education. Psychological disturbance is manifested in drawings in numerous different features which must be interpreted largely on the basis of clinical intuition. One may presume that the college

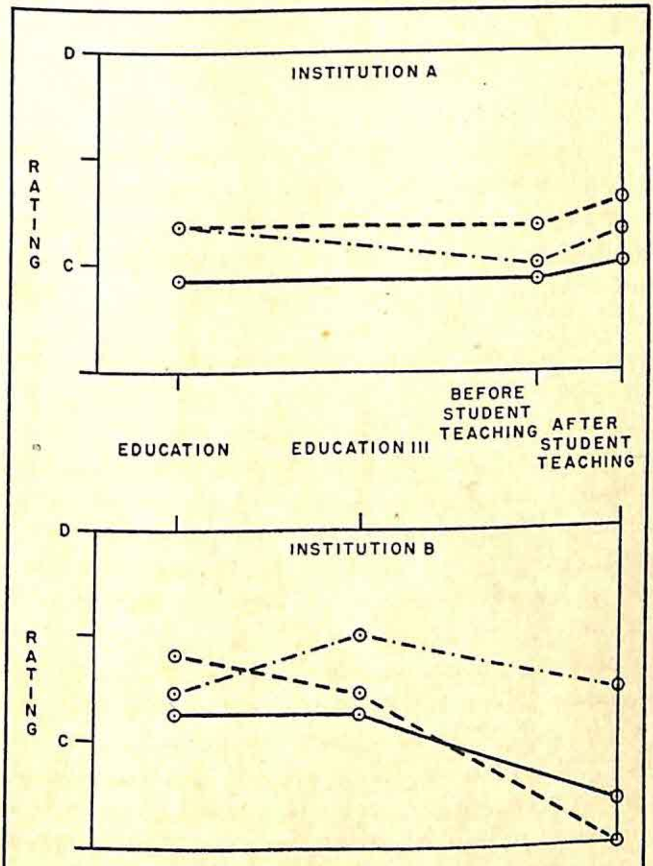


FIGURE 7
AVERAGE RATINGS FOR PSYCHOLOGICAL DISTURBANCE

can exercise some control over the psychological make-up of its students through a program of continuous selection.

It has been pointed out that there are greater opportunities in Institution B than in Institution A for eliminating from the program those who have psychological disturbances which might interfere with classroom teaching. We would thus expect that advanced students of education in Institution B would show a lower incidence of psychological disturbance than corresponding students in Institution A.

In Institution B, advanced students show a slightly lower incidence of psychological disturbance than advanced students in Institution A. These data are summarized in Figure 7, but it is doubtful whether this difference would be reproducible in a new and independent sample.

However, the slight difference in the ratings of the drawings from Institution B at the beginning and at the end of training has no statistical significance although it is in the expected direction. The calculated t for the difference is only 1.41 ($p = 0.2$). Of even less significance is the difference between the drawings produced in the two institutions at the end of training.

SOME CONCLUSIONS

The data presented in this paper demonstrate that one of the variables in the teacher's background; namely, the type of teacher-training program, has marked consequences on the nature of the concepts of teaching that are formed. We are concerned here with variables hypothesized to be partial determinants of how the teacher will eventually perform in the classroom. It must be stressed that we are not dealing directly with how the teacher actually behaves in the classroom. There is obviously a complex relationship between concepts of teaching, other variables, and ultimate teacher behavior in the classroom.

It might be argued that the differences between groups shown by the data are a function of factors other than training conditions. Some differences between groups may be expected because different groups were used at the various levels. However, there are two compelling arguments which tend to refute this argument. First, it may be pointed out that differences between groups at the Education I level are negligible, which implies that prior to training the students in the two institutions are comparable with respect to the variables studied. Second, it must be noted that differences between the education students which appear as training proceeds are entirely in accordance with expectation. Consistency between trends which appear in the groups and the nature of the teacher education program is so great that it is extremely unlikely that these apparent trends could be a product of irrelevant variables resulting from the sampling procedures.

The data also illustrate how some aspects of concepts of teaching may be studied and how, through such study, the results of the teacher education process can be better understood. Drawings represent an admittedly imperfect instrument and will require much improvement before they can be used as anything more than crude research instruments. It seems important to stress, however, that in spite of their crudity, drawings have reflected certain changes in concepts of teaching following teacher-training in different programs. Drawing as a research technique holds great promise and warrants the attention and study necessary to realize its richness. In addition, new instruments are needed in order to supplement the kinds of material which have so far been successfully elicited.

TEAM WORK IN THE COLLEGE CLASS

HENRY CLAY SMITH

Department of Psychology, Michigan State College

The purpose of the present experiment was to test a set of related hypotheses regarding teamwork. Previous studies by the author and his colleagues have explored nondirective and democratic leadership in the college class (*3*, *9*). The former was based upon the conceptions of Rogers (*8*); the latter upon the studies of Lewin and his students (*5*). The present study takes its departure from Nelson's analysis of the integrative leader (*7*, *11*).

The integrative leader sees his primary purpose as the development of teamwork and perceives his authority as resting upon informal group codes. It was assumed in the present study that the success of such a leader would be determined by his skill in applying the principles of integrative leadership. A set of tentative principles was inferred from a consideration of laboratory and field studies of group behavior (*2*, *4*, *5*, *6*, *12*). Specifically it was assumed that the effective integrative leader would: 1) create work for teams; 2) fit the teams together; 3) encourage interaction on the teams; 4) stabilize the teams; 5) reward the teams as well as the individuals; and 6) let the teams decide.

These assumptions concerning the activities of the integrative leader provided the framework from which the set of hypotheses tested in the present study was derived. These hypotheses are numbered below in accordance with the principle to which they are most closely related.

1) Students in a team class are more satisfied and more productive than students in a lecture class.

2A) Students who like to participate are more satisfied and productive in a team class than those who do not like to participate.

2B) Students who are initially better informed about the content of the course are more satisfied and productive in a team class than those who are not.

2C) Students who are low in manifest anxiety are more satisfied and productive in a team class than those who are high in anxiety.

4) The longer students continue as a team, the more productive they become.

5) Students are more satisfied with a grading procedure which rewards group success than with a procedure which rewards only individual success.

6) Students perceive the instructor playing an integrative rôle as warmer, more objective, and more democratic than when he plays the rôle of lecturer.

It was hoped that the simultaneous testing of these hypotheses would clarify the general problem of teamwork as well as its specific applications to the classroom.

METHODS

The general design of the experiment is shown in Table 1. The ninety-six subjects were students in two general psychology classes. Both classes took a ninety-item achievement test and completed a Participation Scale during the first two days of the course. This latter scale was designed to measure student attitudes toward participation in class discussion. On the basis of these test results, the class chosen for the team experiment was subdivided into five groups. These groups eventually named themselves: 'Psychologists Anonymous' (PA), 'Our Gang' (OG), 'Rearguard' (RG), 'Bow Peeps' (BP) and 'Junior Psychologists' (JP). The students were informed that the groups had been equated for initial achievement. However, they were not informed that the groups also varied in their average attitude toward participation from very favorable to very unfavorable.

The forty-eight students in the team class were individually matched on achievement and attitude with one of the sixty students in the lecture group. As a result of this matching, there were individual and group controls within the lecture class. However, no attention was paid to the groups during the lecture class. Students were seated alphabetically and the class procedures were of the informal lecture-discussion type.

TESTS AND SCALES

Achievement tests.—The initial and final tests of achievement consisted of equivalent forms. Each form was composed of ninety multiple choice items: thirty vocabulary items, thirty factual items, and thirty items involving the application of psychological facts and principles. Statistical data on the reliability of initial and final forms as well as for gains scores have been reported previously (*3*).

TABLE 1.—GENERAL EXPERIMENTAL DESIGN

	Team Class (N-48)			Lecture Class (N-48)		
Pretests:	90-item achievement test and participation attitude scale					
	Scores			Scores		
Group by participation attitudes	N	Achievement	Participation	N	Achievement	Participation
Very favorable (P.A.)	10	36.4	39.5	10	36.3	38.3
Favorable (O.G.)	9	37.3	36.8	9	37.0	36.0
Average (R.G.)	10	36.0	33.6	10	35.9	33.6
Unfavorable (B.P.)	9	35.9	30.1	9	35.7	30.4
Very unfavorable (J.P.)	10	36.0	25.8	10	35.9	26.4
Average	48	36.3	33.2	48	36.2	32.9
Methods:	Organized into five teams with group projects, group decisions, and a group incentive			Traditional lecture-discussion class with individual grades as the incentive		
	Both classes had same tests and quizzes. Both completed Anxiety scale.					
End Tests:	Equivalent form of initial achievement test. Scales measuring attitudes toward the course, the incentive system, and the instructor.					

Quizzes composed of similar items were given at the end of the third, sixth, and ninth week of the course. They consisted of forty, fifty, and forty items, respectively.

Participation scale.—The participation scale consisted of the following questions:

1) Do you like to participate in class discussions?
2) Do you find it easy to speak in class?
3) Do you think that class discussions are valuable?
4) Is it easy for you to give an oral report before a class?
5) Do you volunteer answers to questions raised by the instructor?

6) Do you volunteer ideas to start discussions in class?

7) Do you do better in oral quizzes than in written ones?

8) Do you do better than most students when called on unexpectedly?

9) How often have you actually participated in your classes?

10) Do you raise an objection when you disagree with the instructor or a fellow student?

The students were asked to indicate their agreement or disagreement with each question on a five-point scale. Thus the minimum score on the scale was 10, the maximum, 50. The corrected odd-even reliability of the scale was .86.

Course satisfaction scale.—This scale, administered and scored in the same way as the participation scale, consisted of the following five items:

1) The classes in this course have been extremely unpleasant.

2) I found the classes dull and uninteresting.

3) I am dissatisfied with the way this course was run.

4) I never became so involved in the classes that I forgot about everything else.

5) The classes were rather uninteresting.

Incentive system evaluation.—This scale consisted of these items:

1) The grading system used in this course should be discontinued.

2) I did not like the grading system used in this course.

3) I think the grading system used in this course was unfair.

4) I was quite dissatisfied with the method of determining grades.

5) The grading system is not as good as that used in other courses.

Anxiety scale.—The anxiety scale utilized in this experiment was originally developed by Spence and Taylor (*10*). It consists of fifty items which were judged as being indicative of anxiety with fifteen additional items introduced to conceal the anxiety items.

The instructor rating scale.—This scale was constructed to measure the degree to which the students perceived the instructor as warm, objective, and democratic.

It was based on the factor analysis of the dimensions of parent-child relations reported by Baldwin (*1*). It consisted of sixty items adapted from Baldwin's interview rating scales. The first twenty attempted to measure the warm-cold dimension, e.g., "He pays considerable attention to the interests of the students". The second twenty measured the objectivity-emotionality dimension, e.g., "He usually maintains a calm, objective attitude toward the student even in the fact of strong tension". The third twenty measured the democratic-authoritarian dimension, e.g., "He attempts to adjust policies of the class to student wishes wherever possible". The three-dimensions were scored separately. The maximum possible score for each was twenty, the minimum, zero.

CLASS PROCEDURES

In the team class, the five groups were seated in different parts of the room. To encourage group interaction initially, the names of students by groups were mimeographed and passed out and students were given 5 x 8 name cards which each student placed before him during the early meetings.

Grading procedures were also designed to develop group cohesion. An individual grade incentive was used, but the class was informed that each member of the team having the highest average scores on the tests during the course would receive an additional five points on his grade; the next group, four; and so on. Also each team was informed that it would have ten grade points for its group which it could distribute as it decided to the members of the group. Approximately half of their meetings were devoted to projects, designed to further encourage group activity: The determination of reaction time by a group experiment, multiple rôle-playing within the groups, the completion of tests and scales by the group rather than by the individuals, the discussion of case studies, etc. The groups were also permitted to select their own leaders, keep their own attendance records, manage their own class discussions and conduct meetings of their group outside the class. Approximately half of the team class meetings were conducted like those in the lecture-discussion class. A special effort was made to equate the content and method of presentation in these meetings. The lecture class was entirely of the informal lecture-discussion type. The examinations in the two classes were always on the same day.

RESULTS

Tables 2 and 3 present the data used in testing hypotheses 1 and 2A. Table 2 presents the average course satisfaction scores for those groups with the most, intermediate, and least favorable attitudes toward class participation. Table 3 presents the average achieve-

TABLE 2[1].—PARTICIPATION ATTITUDES AND MEAN COURSE SATISFACTION (Total N-96)

Class	Participation Groups			Total
	V.F. (P.A.)	Intermediate (OG, RG, BP)	V.U. (J.P.)	
Team	20.3	16.3	18.4	17.6
Lecture	18.2	17.4	18.3	17.8
Difference	2.1	−1.1	.1	−.2
Difference of Diff	3.2*		1.2	

[1] In this and later tables the following symbols are used:

* 10 per cent level of confidence

† 5 per cent level of confidence

‡ 1 per cent level of confidence

TABLE 3.—PARTICIPATION ATTITUDES AND MEAN ACHIEVEMENT GAINS

Class	Participation Groups			Total
	V.F. (P.A.)	Intermediate (OG, RG, BP)	V.U. (J.P.)	
Team	18.7	27.0	21.1	24.0
Lecture	25.9	21.1	22.6	22.4
Difference	−7.1*	5.9†	−1.5	1.6
Difference of Diff.	13.0‡		7.5	

ment gains in a similar way. The gains were determined by subtracting the score on the initial test from the score on the final.

As indicated in Table 1, the experiment was designed to contrast the effects of a team class and a lecture class. Thus, in these and the following tables, scores in the lecture class are subtracted from those in the team class in an attempt to isolate the influence of the team situation. In Table 2, for example, the average course satisfaction of the forty-eight students in the team class was 17.6, in the lecture class, 17.8. The difference, −.2, indicates that the lecture class was slightly more satisfying than the team class. The comparable figure in Table 3 was 1.6, which indicates the average student in the team class gained 1.6 more points than the average student in the lecture class.

These results lend no support to the hypothesis that team work is more satisfying. They lend little support to the hypothesis that teamwork is more productive. In the light of additional facts the satisfaction results appear somewhat more favorable. Table 7 indicates that the team class was dissatisfied with its group incentive system. Further, the correlation between course satisfaction and satisfaction with the incentive system was .46 in the team class and .22 in the lecture class. It seems quite probable, therefore, that with an equally satisfying grading procedure, the team class would have been more satisfied than the lecture class. In regard to achievement, Table 6 indicates that the achievement superiority of the team class became greater as the course proceeded. Thus, more confidence can be placed in the achievement gains on the total course as one point in a trend than in the one score alone. In any case, the insignificant differences in average results are related to the contradictory trends within the class as indicated by the detailed analyses which follow.

Participation attitudes.—Tables 2 and 3 also present the data which were used in testing hypothesis 2A regarding attitudes toward participation. The hypothesis 2 is obviously not supported by the data. Although the very favorable group (PA) was more satisfied in the team class, their achievement gains were less. Actually the Psychologists Anonymous were the most satisfied but made the least achievement gains of any of the ten groups in the experiment. The intermediate groups also gave contradictory results: Although they reported less satisfaction, they made significantly higher achievement gains than either their controls or the very favorable group. The team group with the least favorable attitude (JP) was most similar in satisfaction and achievement to its control group.

The results in regard to the very favorable group were consistent with the instructor's observations that the group, while talkative, were uncoöperative and individualistic. These results and observations reinforce the results of a previous study (*9*) which showed participants to be more dominating, independent, and nonconforming.

High and low initial achievement.—Hypothesis 2B implies that students who made high scores on the initial achievement test

TABLE 4.—INITIAL ACHIEVEMENT IN RELATION TO COURSE SATISFACTION AND ACHIEVEMENT
(N-60)

Class	Course Satisfaction		Achievement Gains	
	High Init. Ach.	Low Init. Ach.	High Init. Ach.	Low Init. Ach.
Team (N-30)	17.5	17.3	24.3	24.7
Lecture (N-30)	18.5	17.6	24.1	22.3
Difference	−1.0	−.3	.2	2.4
Diff. of Diff.	−.7		−2.2	

would perform better in the team situation than those who made low scores.

Fifteen high and fifteen low achievers in the team class were used to test this hypothesis. They were selected by picking the three highest and the three lowest achievers from each of the five groups. The lecture controls were the individuals originally matched with each of these thirty. The results, summarized in Table 4, do not support the hypothesis. The negative differences are insignificant, but it can be safely concluded that the low achievers were not less satisfied or less productive than the high achievers.

High and low anxiety.—Hypothesis 2C stated that anxious students would be less satisfied and lower in achievement than the unanxious ones. The thirty subjects employed in this check were picked in a similar way to those in Table 4. The results, shown in Table 5, contradict the hypothesis: the anxious students in the team class were somewhat more satisfied and made somewhat larger achievement gains. None of the differences are significant, but there is no evidence that the anxious students were less satisfied or less productive as a result of being in the team situation.

Duration of the Groups.—Hypothesis 4 was based on the assumption that it would take time for the team groups to become efficient. In addition to the initial and final tests, both team and lecture groups took three quizzes during the quarter. The data from these tests were used to test the hypothesis. Table 6 reports the mean scores for the team and lecture groups on the three quizzes and the final test. The results are consistent with the hypothesis: the team

TABLE 5.—ANXIETY IN RELATION TO SATISFACTION AND ACHIEVEMENT (N-60)

Class	Course Satisfaction		Achievement Gains	
	Low Anxiety	High Anxiety	Low Anxiety	High Anxiety
Team (N-30)	18.7	16.6	27.2	24.7
Lecture (N-30)	19.4	15.5	23.1	19.4
Difference	−.7	1.1	4.1	5.3
Diff. of Diff.	−1.8		−1.2	

TABLE 6.—ACHIEVEMENT IN RELATION TO NUMBER OF MEETINGS (N-96)

	Test Results After:			
	12	24	35	44 Meetings
Team	28.5	34.5	32.6	59.8
Lecture	29.1	34.7	32.5	58.2
Difference	−.6	−.2	.1	1.6

class improved from one test to the next. The differences are small but consistent. If it is assumed that the team class would get a lower score on the first test but improve regularly thereafter, the chances of obtaining four differences consistent with this hypothesis would be seven in a hundred.

The incentive system.—Hypothesis 5 stated that students would be more satisfied with a grading system which involved a group incentive than with one which only rewarded the individual. The average differences for the entire group on the right of Table 7 indicate that this hypothesis must be rejected at less than the one per cent level of confidence: students in the lecture class were significantly more satisfied with the individual incentive system than the students in the team class were with the group incentive.

In Table 7 the groups are arranged according to their success under the incentive system. The group (OG) in the team class having the greatest success and getting the greatest reward under the

TABLE 7.—ATTITUDES TOWARD GROUP AND INDIVIDUAL INCENTIVES

	Most Successful (OG-9)	Intermediate Success (RG, BP, PA-29)	Least Successful (JP-10)	Total (48)
Team	17.7	17.9	15.7	17.4
Lecture	19.8	21.8	22.1	21.4
Difference	−2.1	−3.9‡	−6.4†	−4.0‡

system was least different from its control in satisfaction with the system. The intermediate groups (RG, BP, PA) were next. The (JP's) were least successful and least satisfied. The difference between these latter two groups and their controls were statistically significant. Students apparently liked the group incentive system more when they profited by it.

Perception of the instructor.—On the basis of Baldwin's tentative findings (*1*) it was assumed that the most effective instructor is one who is warm, objective, and democratic. However, the study was not designed to test this hypothesis. The instructor attempted to be equally as warm, objective, and democratic in the two classes. The data in Table 8 indicate that his efforts were partially successful: in comparison with the lecture class, the team class rated the instructor as significantly more democratic but less objective and slightly less warm. It seems clear that the same instructor in a team class will be judged as more democratic than when he is in a lecture class.

The perception of the instructor's warmth was related to course satisfaction. The correlation between satisfaction with the course and the rating of warmth of the instructor was .51 for the two groups combined (N-96). The variability in the ratings of warmth of the

TABLE 8.—INSTRUCTOR RATINGS
(N-96)

Class	Warmth	Objectivity	Democracy
Team	15.4	13.9	13.3
Lecture	16.2	14.9	11.7
Difference	−.8	−1.0*	1.6‡

same instructor in the same class was great, varying in the team class from 9 to 20; in the lecture class, from 6 to 20. Differences in anxiety may have accounted for part of this, but not all, since the correlation between satisfaction with the course and anxiety was only -.24 and the correlation between anxiety and warmth, only −.19.

DISCUSSION OF RESULTS

The implications of these results are stated as revised hypotheses with which they are consistent, but do not prove.

1) *Work organized for either individual or group effort is more effectively done by groups.* The results of the present study only suggest that college classes which apply the principles of integrative leadership are more satisfying and productive. The inconclusive findings are not surprising when one considers that the procedures were unusual for both instructor and students. A more definitive test requires not only more experienced instructors but also more suitable materials for such courses. In the meantime, the results are sufficiently encouraging to warrant additional explorations.

2) *Groups organized by sociometric procedures are more satisfying and productive than those which are not so organized.* This hypothesis is consistent with the negative results of the present study regarding achievement and anxiety. Common beliefs about the kind of students that fit best into a team situation were not supported: those high in participation, high in initial achievement, and low in anxiety do not profit more than their opposite. Favorable attitudes toward class participation, the most socially relevant of the variables, was most significantly related to success. However, this relationship was not linear: those who were most favorable to class participation did not do so well as those with less extreme attitudes. It seems probable that group success was more likely to be related to social than to personal variables. Considering the numerous positive findings utilizing sociometric procedures, this hypothesis is advanced with some confidence.

3) *Group incentives based on work standards are more satisfying and effective than group incentives based on competition between groups.* The most confident finding of the present study was that students in the team class did not like the group incentive system. Dislike was related to failure under the system. It seems evident that a team would generally be dissatisfied with a system which

penalizes it. However, the typical group incentives in industry are based on production criteria, not directly on competition between groups. Testing this hypothesis would require the comparison of performance under achievement norms set in advance versus the type of competitive group incentive used in the present study.

4) *A warm objective and democratic leader is more effective than one without these qualities.* Some of the present evidence is consistent with this hypothesis. The team class in which the students rated the instructor as more democratic, made somewhat higher scores. Variations in the rating of warmth of the same instructor were significantly correlated with satisfaction with the course (.51).

SUMMARY

A 'team' class in general psychology with forty-eight students was divided into five subgroups. The subgroups were equated for initial achievement but were varied systematically in their attitudes toward class participation. Each member of the team class was matched on the basis of initial achievement and participation attitude with a student in a lecture class taught by the same instructor.

There was practically no difference in the over-all satisfaction of the two classes. However, the team class made greater achievement gains than the lecture class, and the trend in four different tests was significantly upward. In comparison with the lecture class, the team class was significantly more dissatisfied with their group incentive system; the more unsuccessful the subgroup in the team class, the more dissatisfied it was.

The groups with moderately favorable attitudes toward participation made significantly greater gains than their controls in the lecture class. They also made significantly greater gains than the group with the very favorable attitudes toward participation. Although the very favorable group reported the greatest satisfaction, they made the smallest achievement gains. Compared to their controls in the lecture class, students with low anxiety and high initial achievement scores in the team class were less satisfied and made smaller achievement gains than those with high anxiety and low initial achievement.

REFERENCES

1) A. L. Baldwin, J. Kalhorn, and F. H. Breese. *Patterns of Parent Behavior*. Psychol. Monogr., 58, No. 3.

2) L. Coch, and J. R. P. French, Jr. "Overcoming resistance to change." *Human Relations*, 1948, 1, No. 4, 512–532.

3) D. M. Johnson, and H. C. Smith. *Democratic Leadership in the College Classroom*. Psychol. Monogr., 1953, 67, No. 11.

4) H. J. Leavitt. "Some effects of certain communication patterns on group performance." *J. Abnorm. & Soc. Psych.*, 1951, 46, 38–50.

5) K. Lewin, R. Lippitt, and R. K. White. "Patterns of aggressive behavior in experimentally created social climates." *J. soc. Psych.*, 1939, 10, 271–301.

6) E. Mayo, and G. F. F. Lombard. *Teamwork and Labor Turnover in the Aircraft Industry of Southern California*. Harvard University Business Research Studies, 1944, No. 32.

7) C. W. Nelson. A *New Approach to Leadership: Its Rationale, Data, Reliability, Validity, and Application*. Chicago: Industrial Relations Center, University of Chicago, 1950.

8) C. R. Rogers. *Client-centered Therapy*. Boston: Houghton Mifflin, 1951.

9) H. C. Smith, and D. S. Dunbar. "The personality and achievement of the classroom participant." *J. ed. Psych.*, 1951, 42, 65–84.

10) K. W. Spence and J. Taylor. "Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning." *J. Exp. Psych.*, 1951, 42, 183–188.

11) J. J. Valenti. "Measuring educational leadership attitudes." *J. Appl. Psych.*, 1952, 36, 36–44.

12) C. R. Walker and R. H. Guest. *The Man on the Assembly Line*. Cambridge: Harvard University Press, 1952.

THE RELATIONSHIP OF STYLE DIFFICULTY TO IMMEDIATE RETENTION AND TO ACCEPTABILITY OF TECHNICAL MATERIAL[1]

GEORGE R. KLARE, JAMES E. MABRY, AND LEVARL M. GUSTAFSON

University of Illinois[2]

This study was one of a series[3] on the relationship of various communication variables to the learning of technical training material. 'Style difficulty', as used here, can be defined operationally in terms of the aspects of writing measured by readability and level of abstraction formulas. There are many formulas that attempt to measure the readability of writing, and many reports of their application to general reading materials (see *5* and *6* for a summary). There are, also, many articles suggesting the use of formulas.

Most published studies on the value and validity of formulas have used such measures as readership. Few have used retention or related measures (*6*, *9*), and the one study done on technical material[4] yielded negative results. This study was an attempt to relate style difficulty to immediate retention test score, to amount read in a given time, and to acceptability of technical training material.

---

[1] This research was supported in part by the United States Air Force under Contract No. AF 33(038)-25726, monitored by the Personnel and Training Research Center. Permission is granted for reproduction, translation, publication, use and disposal in whole and in part by or for the United States Government.

[2] Dr. Klare is now at Ohio University; Dr. Mabry at VA Hospital, Salt Lake City, Utah; and Dr. Gustafson at Oklahoma A & M College.

[3] For a more complete presentation of this and other studies in the series, see (*7*).

Grateful acknowledgement is made of the contributions of the following persons to this series of studies: (a) Dr. L. H. Lanier and Dr. L. M. Stolurow, University of Illinois; (b) Mr. E. K. Farson, Mr. F. E. Gregory, Mr. J. H. Orr, Mr. T. E. Carter, Mr. C. S. Hershey, and others, of Chanute Air Force Base, Illinois; and (c) Dr. D. Dubois and Lt. M. Grunzke, Sampson Air Force Base, New York.

[4] "The effect of reading ease on student comprehension of training materials." Hq., Sheppard AFB, 3750 Tech. Trg. Wing. (Mimeo.)

METHOD

Experimental Materials.—A 1201-word multilith-printed lesson from an aircraft mechanics training course was used in this study. The first half of the lesson covered the 'induction system' and the second half the 'cooling system' of the engine. In preparing experimental versions, only the style difficulty, or 'structure', of the material was changed. Technical experts, serving as judges, determined that neither 'content' nor technical terms were changed in any experimental versions developed.

Three 'styles', or levels of style difficulty, were used, 'Easy', 'Present' (the standard version used in the course), and 'Hard'. These styles varied in (a) percentage of short, familiar, frequently used words, (b) average sentence length and the related proportions of simple sentences and prepositional phrases, (c) the proportion of abstract to concrete wordings, and (d) the proportion of active to passive constructions. The 'Easy' style was at about seven to eight grade level and the 'Hard' at about sixteenth plus grade level, as measured by the Dale-Chall readability formula (*1*) and the Flesch readability (*2*) and 'level of abstraction' (*3*) formulas. The 'Present' style was at about eleven to twelve grade level as measured by the Dale-Chall and Flesch readability formulas, and thirteenth to fifteenth grade level by the level of abstraction formula.

Nine 'versions' of the lesson were used in the study. Three were called 'unsplit', treating both halves of the lesson in one style; these were E (for Easy), P (for Present), and H (for Hard). Six versions were called 'split', treating the two halves of the lesson in a different style; these were EP (first half Easy, second half Present), PE, EH, HE, PH, and HP. All versions had the same number of words (1206), and the same format.

Every fifth line of each version was numbered in order to help in getting a measure of amount read by the subjects. Each subject made a tally mark for each complete reading during the experimental period, and indicated the line he was reading when asked to stop.

A fifty-item multiple-choice test was constructed to measure immediate retention. The items were selected from a pool of one hundred and twelve items in such a way that each paragraph of the material was allotted a percentage proportional to its size. The base for writing the items was the Present version of the lesson, so that any bias which might exist would be in the conservative direction (of course, holding content constant in all versions should have tended to minimize bias). The split-half reliability of the test, corrected by the Spearman-Brown formula, was .87.

Acceptability of the styles was determined by the answers to three questions. The first question asked subjects if they noticed a difference in the way the two halves of the lesson were written (that is, in style, not in what was written about, or content). Those who noticed a difference were then asked which half of the material they thought 'easier to read' (question 2) and which half 'more pleasant to read' (question 3). These questions were asked of those subjects who had the unsplit as well as those who had the split versions.

Subjects and Procedure.—Subjects were nine hundred and eighty-nine male airmen in indoctrination training at Sampson Air Force Base (one hundred and nine subjects for PE, and one hundred and ten for all other versions). Various aptitude indices were available for all airmen, based on the Airman Classification Test Battery (these indices, based on a scale of 1 to 9, are conveniently called 'stanines').

Subjects were given twenty minutes to read the lesson material, and during that period indicated the amount read as previously described. Subjects then answered the three questions designed to determine acceptability of the material. After this, forty minutes were allowed for answering the test.

RESULTS

Amount Read (20 Minutes).—Analyses of these data showed that the Hard version resulted in fewer lines read than the Easy or Present versions. The differences, however, were not significant.

The number of words read, on the other hand, differed more. This could occur because, even though all versions had the same number of words, the Easy style occupied fewer lines than the Present, and the Present fewer than the Hard. It was found that the Easy and Present styles resulted in significantly more words read than the Hard, beyond the .01 level. The Easy style resulted in more words read, but not quite significantly more, than the Present. Table 1 presents relevant data for amount read.

Acceptability.—The next analysis involved the three questions asked of all subjects ("did the two halves differ," "which half easier

TABLE 1.—COMPARISONS OF AMOUNT READ IN TWENTY MINUTES FOR THREE LEVELS OF STYLE DIFFICULTY

Version			Means of Reading Measures		
Level	Lines	Words	Times read*	Lines	Words
Easy	126	1206	2.16	272.25	2605.70
Present	134	1206	2.06	276.65	2489.85
Hard	140	1206	1.82	255.36	2307.23

* All subjects read the material at least once.

to read," "which half more pleasant to read"). It was found that, despite instructions to the contrary, subjects' answers were based on content as well as style. This was shown by the fact that the 'cooling system' material (second half of the lesson) was judged easier and more pleasant to read than the 'induction system' material (first half of the lesson) whatever the style combinations used. A critical ratio of 8.77 was found, the probability of chance occurrence being less than .001.

Possible reasons for this include: (a) greater general familiarity with the concept of the 'cooling system' than with the concept of the 'induction system'; (b) greater 'content interest' value of the cooling system; and (c) unpleasant connotations of the term 'induction' for these newly inducted airmen.

The split versions, however, permitted equating for content in order to study the effect of style. Two 6 x 2 contingency tables were set up, split versions x lesson halves. They yielded chi-squares of 21.19 and 29.93 (both significant beyond the .001 level), showing that style had an effect on judgments of "easier to read" and "more pleasant to read" despite the influence of content.

Individual comparisons of the styles were then made, using the formula for the significance of the difference of percentages. Table 2 presents the data relevant to these analyses. It shows that the Easy style was judged significantly easier and more pleasant than the Hard, but not quite significantly easier or more pleasant than the Present. The Present style was judged significantly easier and more pleasant than the Hard. Tetrachoric correlation coefficients between judgments of easier to read and more pleasant to read ranged from .86 to .97.

Immediate Retention Test Score.—Analysis was next made of the

TABLE 2.—PERCENTAGES OF SUBJECTS WHO INDICATED ONE HALF OF LESSON EASIER AND ONE HALF MORE PLEASANT, BASED ON COMPARISONS OF SIX 'SPLIT' STYLE DIFFICULTY VERSIONS

Version*	Percentage Indicating Each Half Easier		Percentage Indicating Each Half More Pleasant	
	1st half	2nd half	1st half	2nd half
EP	39	61	30	70
PE	28	72	22	78
	C. R. = 1.53, P = .06 (point)		C. R. = 1.21, P = .11 (point)	
EH	44	56	45	55
HE	22	78	15	85
	C. R. = 3.07, P < .001 (point)		C. R. = 4.25, P < .0001	
PH	51	49	46	54
HP	34	66	34	66
	C. R. = 2.27, P = .01 (point)		C. R. = 1.62, P = .05 (point)	

* The first letter in each version designation represents the treatment of the first half, and the second letter the second half.

scores on the fifty-item test. A 3 x 1 analysis of variance of scores on the three unsplit versions was computed, followed by individual t-tests of mean score differences. Table 3 presents these data. It shows (a) significant variance attributable to style difficulty, (b) the Easy style resulted in a significantly higher mean score than the Present or Hard styles, and (c) the Present style resulted in a higher (but not quite significantly higher) mean score than the Hard style.

It had been found previously that subjects' mechanical aptitude indices (stanine scores) correlated highly with test scores on the preliminary tests developed for use with the lesson material. Therefore, correlations were computed between mechanical aptitude indices and test scores in this study. The correlations were .71 for the Easy, .75 for the Present, and .74 for the Hard versions. Since there were slight differences between the mean mechanical aptitude indices of the three groups, comparable stratified samples of one hundred Easy, Present and Hard subjects were pulled for analysis. A 3 x 1 analysis of variance and individual t-tests yielded comparable results to those obtained in the previous analysis.

Two further 3 x 3 analyses of variance were possible, utilizing test scores on the split versions. The first analysis involved the use

TABLE 3.—IMMEDIATE RETENTION TEST SCORE DATA, ANALYSIS OF VARIANCE AND *t*-TESTS FOR THREE UNSPLIT STYLE DIFFICULTY VERSIONS

Style Difficulty Version	Number of Subjects	Mean*	SD	Range
Easy	110	27.19	8.68	11–45
Present	110	24.69	8.23	9–45
Hard	110	23.07	7.70	6–48
Source of Variation	**Sum of Squares**	**df**	**Mean Square**	**F**
Between versions	947.03	2	437.52	$6.97, .01 > P > .001$
Within versions	22,209.90	327	67.92	
Versions Compared		***t***		
Easy—Present		$2.25, P = .012$ (point)**		
Easy—Hard		$3.71, P < .001$ (point)		
Present—Hard		$1.46, P = .072$ (point)		

* Maximum possible score, 50; chance score, 12.5.

** These values are based on the use of the normal probability table for greater exactness.

of scores on the first half of the test (which was based on the first half of the lesson material); the second analysis involved scores on the second half of the test. Table 4 presents the results of these analyses.

Table 4 shows that the variance attributable to style difficulty was again significant in both analyses. It also shows that individual mean score differences were comparable to those previously found.

It is possible in this sort of testing that items might be accidentally 'biased' or 'slanted' in favor of a particular version. If such slanting favored the Easy version, for example, mean score differences might be found due to large percentage passing differences on only a few of the total number of items. It will be recalled that using the Present version as the base for the test was an attempt to turn any bias which might exist in the conservative direction. Nevertheless, the subjects who had the Easy, Present and Hard versions were compared as to percentages passing each item as an additional check on bias.

TABLE 4.—ANALYSES OF VARIANCE FOR SIX SPLIT AND THREE UNSPLIT STYLE DIFFICULTY VERSIONS (989 SUBJECTS)*

Source of Variation	Sum of Squares	df	Mean Square	F
		First-half Scores		
Style difficulty	217.82	2	108.91	6.56, $.01 > P > .001$
Position	7.76	2	3.88	.23, $P > .05$
Interaction	130.53	4	32.63	1.97, $P > .05$
Within groups	16,276.15	981	16.59	
		Second-half Scores		
Style difficulty	540.23	2	270.11	10.63, $P < .001$
Position	75.53	2	37.76	1.49, $P > .05$
Interaction	208.33	4	52.08	2.05, $P > .05$
Within groups	24,916.13	981	25.40	

* Version PE had only one hundred nine subjects; in order to keep the N's equal in all versions and thus simplify analysis, the missing subject was assumed to have the mean score of the other subjects.

TABLE 5.—SIGN TESTS OF DIFFERENCES IN NUMBER OF ITEMS IN WHICH SUBJECTS HAVING ONE STYLE DIFFICULTY VERSION EXCELLED THOSE HAVING ANOTHER VERSION

Versions Compared	Number of Items			Critical Ratio
Easy vs. Present	Easy *excelled* 37	No *diff.* 2	Present *excelled* 11	2.82, $P < .01$ (level)
Easy vs. Hard	Easy *excelled* 44	No *diff.* 1	Hard *excelled* 5	5.51, $P < .00001$ (level)
Present vs. Hard	Present *excelled* 38	No *diff.* 2	Hard *excelled* 10	3.95, $P < .001$ (level)

Table 5 presents the data for number of items in which each group excelled. It can be seen that the effects of differences in style difficulty were spread throughout the tests, indicating little evidence of bias.

DISCUSSION

In discussing the results of this study, it might be well to mention briefly certain parts of the study which could not be reported for lack of space. The first involves preliminary work, using the same experimental materials, carried on in the classes of airmen in advanced technical training. It was found that background information on the lesson material, gained in previous training, was very high, and that experimental test score differences between style versions did not appear. Thus, while style difficulty appears to affect immediate retention of subjects who are naïve regarding material, subjects who have considerable knowledge of the material may profit little if any from an easier style of writing.

The second part of the study not mentioned previously in this paper is that regarding delayed retention. Subjects were retested (using the same fifty-item test) two weeks after original testing. Experimental differences, though still in the same direction (i.e., favoring Easy over Present and Present over Hard), had decreased below the five per cent level of significance. It is not known whether this was a function of style difficulty, of the level of difficulty of the test, of the delay between first and second testings (the two-week delay was necessary for administrative reasons; it was longer than desirable), or of other possible variables. An analysis of retention scores did suggest one other possible reason, however. In analyzing second-half test scores, it was found that the style treatment of the first half of the lesson material significantly affected second-half scores. Inspection of the means showed that a Hard first-half treatment resulted in higher mean second-half scores (for Easy, Present and Hard groups combined) than a Present first-half treatment, and Present than Easy. This suggests that beginning with a harder (and less pleasant) style may cause subjects to 'compensate', thus expending more effort on the material that follows. This same tendency, though not significant, was found in the analysis of split-half immediate retention scores (cf. position variance in the two analyses presented in Table 4).

This 'compensation' may be somewhat similar to that found by Laird (*8*) and Freeman (*4*), working with typists. It does seem to run counter to the general principle of 'breaking a subject in' on easy material. However, it should be noted that motivation was very high during the experimental program; lacking high motivation, subjects might not show this behavior. Further work is required, especially since this result suggests that compensation may decrease experimental differences (i.e., between Hard and Easy styles). Certain tentative findings suggest this may have occurred in this study.

CONCLUSIONS

The results of this study indicate that an easier style of writing may result in (a) greater and more complete immediate retention, (b) a greater amount read in a given time, and (c) more 'acceptable' reading material. This was the case for the lesson used in this study, even though content, technical terms, and number of words were held constant.

The study also indicates, however, the great importance of content in determining how well material will be accepted. It shows, in addition, the high relationship between judgments of material as easier to read and more pleasant to read.

REFERENCES

1) E. Dale and Jeanne S. Chall. "A formula for predicting readability: Instructions." *Educ. Res. Bull.* (Ohio State Univer.), 1948, 27, 37–54.

2) R. F. Flesch. "A new readability yardstick." *J. appl. Psychol.*, 1948, 32, 221–233.

3) R. F. Flesch. "Measuring the level of abstraction." *J. appl. Psychol.*, 1950, 34, 384–390.

4) G. L. Freeman. "Changes in tension-pattern and total energy expenditure during adaptation to 'distracting' stimuli." *Amer. J. Psychol.*, 1939, 52, 354–360.

5) S. N. Hotchkiss and D. G. Paterson. "Flesch readability reading list." *Personnel Psychol.*, 1950, 3, 327–344.

6) G. R. Klare and B. R. Buck. *Know Your Reader*. New York: Hermitage House, 1954.

7) G. R. Klare, J. E. Mabry and L. M. Gustafson. "The relationship of verbal communication variables to immediate and delayed retention and to acceptability of technical training materials." Lackland AFB, Tex.: Air Force Pers. & Tng. Res. Center, Dec. 1954. *Research Bulletin.* (AFPTRC-TR-54-103.)

8) D. A. Laird. "Experiments on the physiological costs of noise." *J. Nat. Inst. Industr. Psychol.*, 1929, 4, 251–258.

9) C. E. Swanson and H. G. Fox. "Validity of readability formulas." *J. appl. Psychol.*, 1953, 37, 114–118.

PREDICTION OF THE ACADEMIC SUCCESS OF COLLEGE FRESHMEN

ROBERT A. JACKSON

Michigan State College

All new undergraduates at Michigan State College are required to take four tests: (1) ACE Psychological Examination developed by the Educational Testing Service, (2) Michigan State College Reading Test constructed by the Board of Examiners, (3) Test of English Usage, Third Edition, constructed by Benjamin B. Hickok of the Department of Communication Skills in coöperation with the Board of Examiners, and (4) Arithmetic Proficiency Test constructed by the staff of the Arithmetic Improvement Service. This study deals with the relationship between performance on these tests and academic success as measured by first-term grade point average. The individuals in this study were first-term freshmen at Michigan State College in the fall term of 1952. Only those carrying a full college load (from twelve to nineteen credit hours) were included.

Of the 3053 freshmen tested during Orientation Week, 2983 completed one term of college work; the other seventy students withdrew from college during their first term and received no college grades. Since it was known that the mean grade point average of the women was significantly higher than that of the men, the students were divided on the basis of sex. The means, standard deviations and intercorrelations of the five variables were obtained for the 1687 men and the 1296 women completing one term of college work; they are given in Tables I and II. The grade point average is calculated in the usual manner, with A assigned the value 4 down to F with the value 0; the scores on the four tests are derived scores based upon a ten-point scale calculated from a normal distribution by subdividing the ordinates of the normal curve from $+3\sigma$ to -3σ into ten equal units.

Tests of statistical significance were applied to the data given in Tables I and II and the following conclusions were apparent:

1) MSC freshman men have a significantly higher mean on the Arithmetic Test.

2) MSC freshman women have a significantly higher mean on the English Test and the Reading Test.

TABLE I—MEANS AND STANDARD DEVIATIONS ON THE INDICATED VARIABLES FOR MEN AND WOMEN

Variable	Mean		Standard Deviation	
	Men	Women	Men	Women
Grade Point Average	2.2	2.4	.70	.68
Psychological Examination	5.4	5.3	1.71	1.65
Reading Examination	5.2	5.5	1.70	1.64
English Examination	4.9	6.1	1.64	1.56
Arithmetic Examination	5.6	5.0	1.68	1.61

TABLE II—INTERCORRELATIONS BETWEEN THE INDICATED VARIABLES

(The values above the diagonal are for men, those below are for women)

	Grade Point Average	Psych. Exam.	Reading Exam. (Men)	English Exam.	Arithmetic Exam.
Grade Point Average		.43	.50	.49	.46
Psychological Examination	.52		.74	.56	.47
Reading Examination	.64	.74		.59	.39
English Examination (Women)	.52	.60	.61		.44
Arithmetic Examination	.49	.49	.42	.46	

3) MSC freshman women have a significantly higher mean grade point average.

4) The means of the men and women on the Psychological Test are not significantly different.

5) The best predictor of grade point average is the Reading Test.

Utilization of the multiple regression approach with two or more variables yielded no significantly better prediction of college success than did the Reading Test alone. The preceding data and other knowledge about the four tests indicated the desirability of restricting further investigation to the relationship of the Psychological Test and Reading Test to the grade point average.

The next part of this investigation was concerned with the relationship between grade point average and number of credits carried. The individuals in each of the sex groups were further classified into eight subgroups according to the number of credits carried. The analysis of variance technique was used to test differences in group means. The data are not presented in this report but they indicated that the mean grade point average increased regularly with an increased credit load. The mean score on both the Psychological Test and the Reading Test also increased with credit load. The analysis of covariance technique was next used to adjust grade point averages on the basis of Reading Test scores.[1] The adjusted mean grade point averages were still significantly different but at a much lower level of significance than the unadjusted grade point average means. This indicated one of the major causes of differences in mean grade point average to be differences in reading ability; however, other factors not included in this study are operating in such a manner as to cause the adjusted mean grade point average to increase with credit load. In other words, students taking heavier credit loads are generally more able students and continue to perform at the expected higher level despite the heavier load.

The mean scores on the Psychological and Reading Tests for those men completing one term and for those withdrawing during the first term are given in Table III; similar data for the women are given in Table IV. From Table III it is apparent that the men who withdrew during the first quarter had lower mean scores on both of the tests than the group completing one term. Table IV

TABLE III—MEAN SCORE ON THE PSYCHOLOGICAL AND READING TESTS FOR THE INDICATED GROUPS OF MEN

	N	Psychological Exam.	Reading Exam.
Men withdrawing during first term	48	4.56	4.38
Men completing first term	1687	5.38	5.21

TABLE IV—MEAN SCORE ON THE PSYCHOLOGICAL AND READING TESTS FOR THE INDICATED GROUPS OF WOMEN

	N	Psychological Exam.	Reading Exam.
Women withdrawing during first term	22	5.14	5.59
Women completing first term	1296	5.33	5.54

TABLE V—PERCENTAGE OF EACH ABILITY GROUP IN INDICATED GRADE POINT GROUP (MEN)

Ability on Psychological Examination	Grade Point		Ability on Reading Examination	Grade Point	
	Passing	Failing		Passing	Failing
High	94.6	5.4	High	96.6	3.4
Middle	78.7	21.3	Middle	80.2	19.8
Low	51.6	48.4	Low	48.8	51.2

TABLE VI—PERCENTAGE OF EACH ABILITY GROUP IN INDICATED GRADE POINT GROUP (WOMEN)

Ability on Psychological Examination	Grade Point		Ability on Reading Examination	Grade Point	
	Passing	Failing		Passing	Failing
High	97.5	2.5	High	98.6	1.4
Middle	86.6	13.4	Middle	87.0	13.0
Low	55.9	44.1	Low	48.7	51.3

shows that the women withdrawing from school during the first term had mean Reading and Psychological Test scores nearly identical with the mean scores of those completing one term. Apparently factors other than ability are operative in the case of women withdrawing from college. Of the freshmen tested during Orientation Week, 2.8 per cent of the men and 1.7 per cent of the women withdrew during their first quarter.

Correlations between test scores and grade point averages have already been given, but these are rather difficult to use directly. A second method of analyzing the relationship between grade point average and test scores involves the computation of the probability that persons with a given score on the predictor will achieve various levels of success on the criterion. This method helps to communicate to individuals the probability of doing at least as well as certain cut off points on the criterion. Since the critical grade point is a C average, it seemed desirable to determine the chances of various students doing at least average work in their first term of college. It was felt that many extreme situations in the first quarter of college might tend to result in the first-term grade point average being unrepresentative of the individual's eventual grade point average; therefore, the grade point distribution was dichotomized at a grade point average of 1.75. Those individuals with a grade point average of 1.75 or above were called the passing group and those with a grade point average less than 1.75, the failing group.

To obtain percentages for the extreme groups on the basis of larger numbers, the ten-point scale was condensed into a three-point scale: (1) high-ability group (derived scores 10, 9, 8), (2) middle-ability group (derived scores 7, 6, 5, 4), and (3) low-ability group (derived scores 3, 2, 1). Two tables were set up with the dichotomized grade point forming one classification and these three ability groups the other. The entries in the tables are the per cent of each ability group that were in the passing or the failing group. The data for the men are given in Table V and for the women in Table VI.

Tables V and VI indicate that a much greater percentage of the low-ability group will obtain less than a C average than either of the other two groups. Practically all of the high-ability group do obtain an approximate C average or better. About fifty per cent of those individuals with a derived score of 3 or less will fail to make a C average the first term in college.

In interpreting these statistical data, the following limitations should be kept in mind:

1) A certain error of measurement pertains to each derived score; thus, the true score for a particular individual on one of the tests might be 6 rather than his observed score of 5, or 4 rather than 5.

2) The grade point average is based on one quarter of work; thus a particular student's grade point average may not be representative of the total grade point average he will eventually compile, due to possible extreme circumstances in this, his first quarter in college. For example, a student with high ability may do relatively ineffective work in certain areas due to lack of interest, motivation, etc.

3) The probabilities presented are based on group performance, and do not necessarily hold for individual prediction.

After considering these limitations, however, it seems that several statements are appropriate:

1) Women obtain significantly higher grades than men.

2) The Reading Test is the best predictor of academic success.

3) Women tend to perform more nearly in accord with their measured ability than do the men.

4) In general, the percentage of an ability group failing to obtain at least a C average increases as the ability of the group decreases, but some students in the low-ability group make satisfactory grade point averages.

5) Practically all of the individuals in the high-ability group obtain at least a C average; whereas approximately fifty per cent of those in the low-ability group fail to obtain a C average.

A STUDY OF THE EFFECTS OF TEACHER KNOWLEDGE OF PUPIL CHARACTERISTICS ON PUPIL ACHIEVEMENT AND ATTITUDES TOWARDS CLASSWORK[1]

KENNETH B. HOYT

College of Education, State University of Iowa

To say that individual differences exist in school children has little significance simply as an observation. The significance lies in the actions one can take as a result of having made this observation and the results which ensue in terms of changes in students.

Many people prominent in education have, over the last fifty years, urged teachers to discover the ways in which and the degrees to which school children differ in various characteristics (*1*, *2*). Some have been so bold as to state specifically such beliefs as "There is no doubt that better teaching results from supplying teachers with pertinent facts about all their pupils." (*6*).

THE PROBLEM

One approach to the problem of the relative advisability or inadvisability of providing teachers with information concerning pupil characteristics is to study some of the effects resulting from such knowledge on the part of teachers. This investigation was concerned with studying the effects of teacher knowledge of pupil characteristics on pupil achievement and teacher-pupil relationships. Seven major questions were devised which served as the focus for the investigation. These were as follows:

1) What is the effect of providing teachers with systematic appraisal data relative to pupil characteristics on teacher knowledge of pupil characteristics?

2) What is the effect of teacher knowledge of pupil characteristics on pupil achievement?

---

[1] Summary of Ph.D dissertation presented to the Graduate Faculty at the University of Minnesota. The author is pleased to acknowledge the advice and encouragement which Professor Willis E. Dugan gave to this study.

3) Does the effect of teacher knowledge on pupil achievement act differentially for boys and girls?

4) Does the effect of teacher knowledge on pupil achievement act differentially on students from different learning ability levels?

5) What is the effect of teacher knowledge of pupil characteristics on pupil attitudes towards teachers?

6) Does the effect of teacher knowledge on pupil attitudes towards teachers act differentially for boys and girls?

7) Does the effect of teacher knowledge on pupil attitudes towards teachers act differentially on students from different learning ability levels?

The major experimental factors studied with reference to the treatment effect were teacher knowledge, sex, and ability level. The major experimental variables were pupil achievement and pupil attitudes toward teachers.

POPULATIONS STUDIED AND SAMPLING METHODS EMPLOYED

The experiment was conducted during the school year 1953–54 utilizing eighth-grade students in English, mathematics, and social studies classes from two public junior high schools in Minnesota. In School A, two hundred and ninety students were included on the roster and in School B, three hundred and fifty-six students. The actual experimental period was six months in length and all findings should be interpreted with this limitation in mind.

Individual students in each school were selected for and assigned to experimental sections by the method of random sampling within sex and scholastic aptitude sub-populations. Six sub-populations were formed from the roster in each school with boys and girls each being divided into highest, middle, and lowest ability groupings on the basis of obtained IQ scores on the *Otis Quick Scoring Mental Ability Test*, Beta level, Form Cm, which had been administered to all students in both schools during the school year 1952-53.

In School A, ninety students were selected and assigned at random, thirty to each of three experimental sections. Complete analysis data were available for eighty-five of these ninety students at the conclusion of the experimental period. In School B, one hundred and twenty students were selected and assigned at random, forty to each of three experimental sections, with complete anal-

Table 1—Numbers of Students, by Sections, in Each Stratum for Which Complete Analysis Data Were Available in School A

	Section 8A		Section 8B		Section 8C		Totals	
	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys
High IQ	5	5	5	4	5	5	15	14
Middle IQ	4	5	5	4	6	4	15	13
Low IQ	5	5	5	4	5	4	15	13
Totals	14	15	15	12	16	13	45	40

Table 2—Numbers of Students, by Sections, in Each Stratum for Which Complete Analysis Data Were Available in School B

	Section 8D		Section 8E		Section 8F		Totals	
	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys
High IQ	7	5	7	5	7	6	21	16
Middle IQ	5	8	6	6	7	6	18	20
Low IQ	6	6	6	6	5	5	17	17
Totals	18	19	19	17	19	17	56	53

ysis data being available for one hundred and nine students. The effects of the loss of these students as well as a graphical representation of the sub-populations involved can be obtained from an examination of Tables 1 and 2.

SPECIFICATION AND ASSIGNMENT OF THE TREATMENT VARIABLES

Three distinct treatments were applied to each section during the experimental period. These were arranged for both schools in such a way that any one teacher was responsible for a different treatment with each section. Definitions of the three treatment variables are presented in terms of the amount of information concerning pupil characteristics given to teachers.

The treatment of 'No Information' (N) is defined as the process of giving the teacher only the names of the pupils and no other systematic information.

The treatment of 'Test Scores' (T) is defined as the process of giving the teacher the pupils' test scores. These consisted of converted IQ scores from the *Otis Quick Scoring Mental Ability Test*, Beta level, Form Cm. (which all students took during the school year 1952-53); scaled scores, by parts, for the *Coöperative English Test*, Lower Level, Form Z; the *Coöperative Social Studies Test for Grades 7, 8, and 9*, Form X; and the *Coöperative Mathematics Test for Grades 7, 8, and 9*, Form X (which each student took in September, 1953.) No other systematic information was given the teacher.

The treatment of 'Tests Plus Other Information' (TO) is defined as the process of giving the teacher, for each student, all the information described for the Test Scores treatment plus a pupil information blank which was administered in September, 1953, and constructed by the experimenter. In addition, each teacher was encouraged to discover all he possibly could about the characteristics of each student receiving this treatment.

These treatments were assigned to experimental sections by subject using a table of random numbers compiled by Fisher and Yates (*3*). The resulting assignments for School A are shown in Table 3 and for School B in Table 4.

TABLE 3—ASSIGNMENT OF TREATMENT VARIABLES TO EXPERIMENTAL SECTIONS IN SCHOOL A

Section	Teacher 1 (English)	Teacher 2 (Mathematics)	Teacher 3 (Social Studies)
8A	TO	N	T
8B	T	TO	N
8C	N	T	TO

TABLE 4—ASSIGNMENT OF TREATMENT VARIABLES TO EXPERIMENTAL SECTIONS IN SCHOOL B

Section	Teacher 4 (English)	Teacher 5 (Mathematics)	Teacher 6 (Social Studies)
8D	T	TO	N
8E	TO	N	T
8F	N	T	TO

DESCRIPTION OF TESTS AND FORMS USED

Student achievement was measured in September and again in April using the *Coöperative English Test*, Lower Level, Form Z, the *Coöperative Mathematics Test for Grades 7, 8, and 9*, Form X, and the *Coöperative Social Studies Test for Grades 7, 8, and 9*, Form X. While these tests measure recall and recognition, it is apparent that what some would regard as the relatively more important goals of retention and transferability of learned materials to new situations is not measured by these instruments. This serves to specify one limitation of the study.

Information relative to other pupil characteristics which were considered as important aspects of treatment TO were collected in September by administration of the pupil information blank.

Student attitudes towards particular teachers were measured by use of an inventory constructed by the experimenter. It consisted of items taken from the *Student Reaction Inventory* (*4*), contained in category 7, "Individualization of Instruction," which, on an item analysis, showed differences between 'good' and 'poor' teachers (5) significant at the 1% level. Forty such items were selected and five more non-discriminating items which had high face validity were added making forty-five items in all. Responses typical of those students make about 'good' teachers were assigned a weight +1; those associated with 'poor' teachers a weight of −1; and all items answered 'undecided' were scored as zero. This instrument was administered to each student three times, once in each of the three experimental groups to which he was assigned.

Each teacher was asked to complete a blank consisting of questions designed to measure the degree to which he was able to recall specific characteristics of certain randomly selected pupils in the various experimental sections. Six students, one from each stratum, were picked at random from each experimental section in a given school and the three teachers involved were asked to complete the blanks for these eighteen students.

EXPERIMENTAL PROCEDURES

The experimental procedures employed in this investigation can perhaps be best described in terms of controls applied during the experimental period to the various treatment procedures. Rigid controls were also applied to the collection of data, but these are not described here.

To maintain the limitations prescribed by treatment N teachers were urged, for each member of the section where this treatment was applied, to: (a) Refrain from examination of the pupil's cumulative record folder; (b) Refrain from discussion of this pupil in conversations with other school personnel; (c) Refrain from discussing with the student any topics which might lead to information concerning his activities outside of the classroom; and (d) Refrain from seeking out parents of the student for consultation and discussion.

To carry through the procedures prescribed by treatment T teachers were urged, for each member of the section where this treatment was applied, to maintain the same limitations described for treatment N. They were given copies of all initial achievement test scores and IQ scores for individual students in this section. In addition, graphs of distributions of these scores were supplied as was a scattergram showing the relationships between intelligence and achievement test scores in the subject the teacher was teaching.

To carry through the procedures prescribed by treatment TO the teachers were supplied, for each member of the section where this treatment was applied, with all of the information described for the T treatment. In addition, they were given: (a) Individual record folders for each student in which they were encouraged to place records of all information learned regarding characteristics of the student; (b) The pupil information blanks which the student had completed and on which the experimenter had marked some interpretive suggestions; and (c) A four-page group analysis of the section made on the basis of information obtained from the pupil information blanks.

To work towards a clear differentiation of treatments individual conferences were held with teachers at least once every other week during the experimental period. Here, in addition to providing frequent reminders of the restrictions imposed by Treatments N and T, the experimenter worked with each teacher in interpreting the significance of the data compiled for students in the T and TO sections. No attempt was made to force the teacher to discuss every pupil; rather, the choice of pupils to be discussed was always left with the teacher.

To work towards the prevention of gross changes in teaching method teachers were asked to: (a) Use the same textual material for all experimental sections; (b) Cover the same units at approximately the same time with all experimental sections; (c) Use the same general approach in teaching method with all experimental sections, i. e., if a topic were presented by means of a student panel discussion in one section, this same approach should be used in all sections. Checks made during and at the conclusion of the experimental period satisfied the investigator that these restrictions were being carried out.

STATISTICAL DESIGN AND METHODS OF ANALYSIS

The statistical design utilized for analysis of achievement test scores is defined as 2 × 2 × 3 × 3 factorial. Specific factors involved included two schools, two sexes, three ability levels, and three treatments. This design was applied three times, once for each of the three school subjects—English, social studies, and mathematics. The method of statistical analysis employed was the analysis of variance and covariance with unequal frequencies in the subclasses using unweighted means. The method of unweighted means rather than the least squares method was chosen primarily because of the interest of the investigator in examining first-order interaction effects of schools, sex, and ability with the Treatment effects in the experimental design. Where analysis of covariance was employed, adjustment was made on the basis of initial achievement test scores.

The statistical design utilized for analysis of attitude test scores is defined as a 2 × 3 × 3 × 3 factorial. This design was applied twice, once for School A and once for School B. Specific factors included two sexes, three school subjects, three ability levels, and three treatments. The method of statistical analysis employed was the analysis of variance with unequal frequencies in the subclasses using unweighted means.

The statistical design utilized for analysis of teacher knowledge scores is defined as a 3 × 3 factorial with three teacher subjects and three treatments representing the factors involved. This design was applied twice, once for School A and once for School B. The method of statistical analysis employed was the analysis of variance with equal frequencies in the subclasses.

The assumptions involved in the use of these various statistical tools were tested separately for each design and found to be satisfied.

SUMMARY OF FINDINGS

1) Teacher knowledge of pupil characteristics was increased through the use of appraisal data. Significant F ratios were obtained on this analysis for mean treatment differences, for mean teacher-subject differences, and for mean interaction differences of treatment with the various teacher-subject combinations. The mean square for the treatment effect provided a significant F ratio when compared with the mean squares for error, teacher-subject, and interaction thus establishing itself as probably constituting a true effect. This finding indicates that the Treatment procedure did operate as intended.

2) Teacher knowledge of pupil characteristics was not found by itself to result in increases in pupil achievement in mathematics, social studies, or English. The application of the various treatment effects did not produce any significant differences in adjusted mean achievement test scores in mathematics and social studies. While significant adjusted mean differences in favor of Treatment N were found for adjusted achievement test scores in English, the complexities of the interaction effects present prevent one from regarding the observed differences as true treatment effects.

3) Teacher knowledge of pupil characteristics was not found to act preferentially on either sex in terms of its effect on increases in pupil achievement in mathematics, social studies, or English. There were no significant F ratios obtained for the interactions of the sex with the treatment factor.

4) Teacher knowledge of pupil characteristics was not found to act preferentially on ability levels of pupils in terms of its effect on increases in pupil achievement in mathematics, social studies, or English. The interaction effects of ability level with the treatment factor were not significant in mathematics and social studies classes. A significant F ratio was found for the interaction between ability level and treatment in English classes. However, because of significant interactions also present between treatment and schools and between ability and schools plus significant main effects for school and treatments, the exact interpretation of the significant interaction of ability with treatment was not clear.

5) There was a definite tendency for increases in teacher knowledge of pupil characteristics to improve pupil attitudes towards teachers. Significant differences in favor of Treatment TO were found in School B. In School A, the F ratio of 1.461 was smal[illegible]

than that of 3.04 required to reject the hypothesis of equality at the 5% level of significance. However, the mean of 8.37 for Treatment N, 8.77 for Treatment T, and 12.12 for Treatment TO were all in the direction indicated by the significant findings in favor of Treatment TO which were found in School B.

6) Teacher knowledge of pupil characteristics was not found to act preferentially on either sex in terms of its effect on pupil attitudes toward teachers. The interaction of sex with treatment was not significant in either school.

7) Teacher knowledge of pupil characteristics was not found to act preferentially on ability levels of pupils in terms of its effect on pupil attitudes towards teachers. The interaction of ability level with treatment was not significant in either school.

REFERENCES

1) Warren, Baller, *et al.* "Essential pupil-personnel records." *Teachers College Record*, January, 1947, 268–86.

2) R. N. Bush. "A study of student teacher relationships." *Journal of Educational Research*, May, 1942, 645–656.

3) R. A. Fisher and Frank Yates. *Statistical Tables for Biological, Agricultural and Medical Research*. New York: Hafner Publishing Company, Inc. 1949.

4) Paul R. Grim and Cyril J. Hoyt. "Appraisal of teaching competency" *Educational Research Bulletin*, April 16, 1952, 85–91.

5) ———. "Excerpts from two instruments for appraising teaching competency." *Journal of Educational Research*, May, 1953, 705–710.

6) Ralph D. Wadsworth. "Give teachers the facts about pupils." *California Journal of Secondary Education*, February, 1946, 84–87.

BOOK REVIEWS

John Dollard, Frank Auld, Jr., Alice Marsden White. *Steps in Psychotherapy*, New York, The Macmillan Company, 1953 pp. 222. $3.50

Harrington V. Ingham and Leonore R. Love. The *Process of Psychotherapy*, New York, McGraw-Hill Book Company, Inc., 1954, pp. 270. $5.00

O. Hobart Mowrer and Twenty-One Contributors. *Psychotherapy: Theory and Research*, New York, The Ronald Press Company, 1953, pp. 700.

Carl A. Whitaker and Thomas P. Malone. *The Roots of Psychotherapy*, New York, The Blakiston Company, 1953, pp. 236. $4.50

Lewis R. Wolberg. *The Technique of Psychotherapy*, New York, Grune and Stratton, 1954, pp. 869. $14.75

It might be appropriately asked why five books on psychotherapy should be reviewed in The Journal of Educational Psychology. There are many reasons for this. In the first place, psychotherapy has been described over and over again by its protagonists as a process of learning. May, for instance, in Mowrer's book states: "The historical and logical analysis of the preceding pages indicates that psychological therapy is a learning, or more accurately, a relearning process." (p. 33) While Shoben in the same book states: "Psychotherapy may be fruitfully understood as a learning process." (p. 125) The cover of THE JOURNAL OF EDUCATIONAL PSYCHOLOGY specifically states that the journal is devoted primarily to the scientific study of problems of learning. Psychotherapy has also been defined as personality change and the Journal is interested in the development of personality. The second reason for reviewing books on psychotherapy in this Journal is that the principles of transference, countertransference and resistance apply equally to education as well as to psychotherapy. Children bring to the classroom situation attitudes toward the teacher and toward their classmates from their earlier experiences as incorporated in their personality trends which can be called transference attitudes. And teachers likewise bring to the classroom situation attitudes out of their own past which can be called countertransference. Children exhibit many kinds of resistance to learning and it is p sible that the educator can profit from an understanding of some of the principles of the process of psychotherapy insofar as he can reinterpret these principles in terms of the educational process. In the third place, psychotherapy is coming more and more to be applied by college counselors, school psychologists, and even teachers to the problems of adjustment that their students present. In the fourth place, psychotherapists are beginning to find that the psychological tests used by the clinical psychologist are of some value to them—not only in order to assist them in making their preliminary diagnosis of the client's difficulties, but also, as Bellak has pointed out, as an integral part of the psychotherapeutic process itself. In the fifth place, many psychotherapists are now using methods that are essentially educational and the educator can perhaps profit from the experience of therapists with these methods.

The book by Dollard, Auld, and White might be called a supplement to the earlier book by Dollard and Miller, entitled *Personality and Psychotherapy*.[1] This new book contains the report of one case of Mrs. 'B', which serves to illustrate the principles and practice of psychotherapy as they were enunciated more formally and abstractly in the earlier book by Dollard and Miller. The process of psychotherapy which this case presents is a strange combination of nondirectiveness and directiveness. The patient is almost ordered to free associate, and the authors even recommend the use of mild threats in order to steer their client's responses along the lines of free association. In the third interview, we are told: "The therapist observed that with this hour the patient would start free association. Mrs. 'B', replied, 'I do not feel like talking'." Because of this refusal to begin free associating, Dollard in his supervisory comment suggests saying the following: " 'You have just a little time here. You have to decide how to make use of it. I think you are in sufficient trouble so that you should seriously consider keeping your appointments here'! You must motivate her to come and remind her of the misery that she is bearing." (p. 41)

Dollard's therapeutic process might be called an aggressive form of nondirectiveness. As supervisor, he instructs his therapists to proceed without hesitation in his tasks of labeling and interpreting. For instance, he suggests the following: "The ideal procedure in therapy would be first to establish the fact of the conflict, then to recur to its origin. Being able to label the feeling and behavior appropriately could have a great effect." (pp. 109, 110) Apparently Dollard believes that the therapist should do this labeling and interpreting with assurance and that the therapist should expect to do most of the analytic work himself.

---

[1] John Dollard and N. E. Miller. *Personality and Psychotherapy*. New York: McGraw-Hill Book Co., 1950.

In reporting the case of Mrs. 'B' the upper part of each page is a report of the therapist who conducted the case, and the lower part of the page contains the suggestions made by the supervisor. This enables the reader not only to follow the treatment of the case, but also the supervisor's explanatory and interpretive comments with regard to the therapist's approach.

Part II of the book gives the results of psychological tests given to Mrs. 'B' and comments on their value. The authors come to the conclusion: "For our own part, we do not believe that the test reports would have been very helpful." (p. 197)

The book by Ingham and Love is an excellent example of teamwork between a psychiatrist and a clinical psychologist. The senior author is a psychiatrist attached to the student health service of the University of California at Los Angeles, and the junior author is a clinical psychologist on the staff of the Los Angeles Psychiatric service. No indication is given as to the relative contribution of these authors to the text. There is good reason to believe, however, that Miss Love conducted the second of the two cases presented in the appendix. This is an elementary and readable text on the process of psychotherapy. One gains the impression that the authors are insightful persons and that they are sensitive to the relationships between a therapist and his client. Technical vocabulary is avoided and the book is self-contained with no bibliography or footnote references to the work of other individuals.

The book by Whitaker and Malone on the other hand is much more of a scholarly treatise. The first author is probably the practitioner and is interested in process and techniques. The second author is more the scientist and is concerned with the relations between psychotherapy and biology. The book itself uses a somewhat unfamiliar vocabulary; and in spite of the fact that there is a glossary of terms at the end it is difficult reading because even familiar terms are used with a somewhat unfamiliar meaning. These authors are in the business of training psychotherapists at the Emory University School of Medicine, and the report is in part based on their own innovations. Whereas the book by Ingham and Love is addressed primarily to therapists dealing with the milder problems such as would be found in a college population, Whitaker and Malone are dealing with the problems of deeply regressed, probably schizophrenic individuals; and, consequently, psychotherapy as they envision it departs much more from ordinary counseling. One of the authors has made a distinct contribution in showing how personality as well as the physical body can be depended upon to have its own process of repair or readaptation, and that it is this process of seeking better adaptation and adjustment that every psychotherapist must depend upon as the inherent healing process.

One of the major contributions of this book has to do with new developments in the theory of countertransference. More than any other writer they show how the therapist brings his own problems into the psychotherapeutic situation. They introduce the interesting concept of the therapist becoming the patient on occasion when his own problems become intense and the patient reversing the rôle and serving as a therapist to the therapist. They also introduce the concept of the 'sliver problem'. These slivers represent areas of minor transference difficulties which have not been worked through in the therapist's own psychotherapy. "In the ordinary person, these are unimportant interpersonally since they do not involve critical areas. To the professional therapist, however, because of the multiplicity of his therapeutic relationships, they give rise to frequent and varied interpersonal *impasses*. During therapy these are continually being worked through by the therapist-vector in the patient, and they represent a major source of satisfaction to the therapist. Each time an *impasse* is resolved, such a sliver problem has probably been resolved in the process." (p. 151) These authors point out how the practice of psychotherapy contributes to the therapist's own readjustment.

The book edited by Mowrer is difficult to cover in the small space which can be allotted to it in this review. All of the contributors are clinical psychologists and all but two of them have the Ph.D. degree. One is impressed, however, by the great diversity of approaches and points of view made by these clinical psychologists; an outside observer must be impressed by the fact that psychotherapy means many different things to clinical psychologists. In the first chapter, May emphasizes the point that it is impossible for a therapist to avoid introducing his own personal values into the therapeutic situation. This point of view is contrary to that held by Rogers in Chapter 2 who, as is well known, is an out-and-out nondirectivist.

Mowrer has taken opportunity to elaborate further his theories concerning neurosis in five of the chapters of the book of which he is the editor. This reviewer has searched diligently in order to find out exactly what psychotherapy means to Mowrer; and he finds it a little discouraging to read: "Psychotherapy is a complex procedure and there are many valid and suggestive ways of describing it. . . . No attempt will be made to describe this procedure comprehensively." (p. 148) One would expect something better than this in a volume of seven hundred pages. It is clear that Mowrer believes "that psychotherapy is not that of emotionally re-educating the patient, but of helping establish problem-solving habits which will enable emotions to operate as they are normally intended to." (p. 149) Drawing on his two-factor theory of learning, Mowrer states: "Although the neurotic's principal complaint is that of emotional suffering, the most effective therapeutic attack is made in the area of problem-solving behavior." (p. 148) One might conclude from this statement that Mowrer's psychotherapeutic approach is essentially directive, that is, in helping his patient forward in his problem-solving behavior. However, Mowrer contradicts this conclusion by stating: "Attempts to deal with the neurotic's problems either directly or directively have not brought very satisfactory results, and we can now see that in some ways these approaches probably perpetuate the strategies which constitute the very essence of neurosis." (p. 148) Elsewhere, Mowrer describes therapy as follows: "The person is invited to *talk*. . . . With an appropriate degree of tentativeness and timeliness, the therapist then begins to report *his* associations to the patient; and it is these which constitute the essence of 'interpretation' and which set in motion forces which are designed to arrest and reverse the dissociative trends within the patient." (p. 79) "The associations of the therapist serve to reinstate the conflicts which the repression resolved." (p. 79). "Instead of seeking to advance the patient's psychological welfare by dissolving a supposedly too severe superego, a therapist operating within the present frame of reference *accepts* the patient's high standards as evidence of good character and works toward the goal of making these standards more understandable and helpful, so that they may ultimately be harmoniously integrated with the rest of the personality. Consistent with this point of view, the therapist also takes the attitude that depression, inferiority feeling, anxiety, and the other painful affects which the neurotic experiences are *signs of strength* within the personality." (p. 91)

In spite of the fact that Mowrer believes that the task of psychotherapy is not that of emotionally re-educating the patient, he has included a chapter (19) on physiological changes during psychotherapy by O'Kelly. This would indicate that Mowrer either is very broadminded and is willing to include all points of view, or he himself failed to recognize the inconsistency. Part II of this book is a collection of papers dealing with research methods and results that have bearing on psychotherapy. The research-minded psychologist will find a mine of information in these chapters dealing with research procedures currently being experimented with in testing some of the outcomes of psychotherapy.

The massive volume by Wolberg stands apart from these others. For the first time we have a comprehensive manual dealing with psychotherapy in its many aspects. This book undoubtedly will serve a as text in institutes for the training of psychotherapists and as a manual and reference book by established practitioners. Part I of the book deals with the general principles of psychotherapy. Wolberg divides therapy into three main types: supportive therapy, re-educative therapy, and reconstructive therapy. The first of these is essentially environmental manipulation; the second is what is commonly known as counseling; whereas the third is depth therapy which results in personality reorganization. Three chapters (5, 6, and 7), describing the variations under each of these headings constitutes a remarkably helpful survey of the different schools of psychotherapeutic thought, and the student will find Chapter 7 of value in orienting him to the different schools of psychoanalysis and neo-Freudianism. Parts II, III, and IV of the book deal with the beginning, middle and terminal phases of treatment. Of particular value are the copious case materials which illustrate adequately the theoretical points made in the text. In Part V special aspects are dealt with including the handling of emergencies, the problem of supervision, a chapter answering questions which therapists frequently ask, and the problems of recording. Chapter 50 dealing with psychotherapy in special conditions stands apart from the others. This reviewer would not find it difficult to believe that this chapter was written before the rest of the book for it represents [illegible] point of view about psychotherapy which is not consonant with the main parts of the book since it represents an older and more strictly old-fashioned psychiatric approach. For instance, in the section on the treatment of alcoholism we read: "Sometimes a belittling and challenging manner mobilizes in the patient a need to prove that he has the 'guts' to master the craving for alcohol." Wolberg adds, however: "This, however, cannot be depended upon." (p. 616)

In spite of the fact that Wolberg is eclectic in this book and gives fair treatment to many points of view, his own approach to psychotherapy is essentially that of encouraging his patient to think through the causes for his difficulty. Certainly in the case illustrations Wolberg is more active than one would expect to find a psychoanalyst. For instance, one of his patients says: "Yes, I see what you mean doctor. If I can understand why I feel the way I do everything will be all right?" (p. 701) Then later Wolberg responds to this patient: "The only way you can change is to find out how the patterns you have been accepting as normal are abnormal." (p. 753) Apparently, therefore, Wolberg puts the entire emphasis in his therapeutic work on understanding and insight with the belief that when these are achieved the patient has overcome the greatest obstacles toward his recovery. In his activity he actually goes so far as to press his patients to make their own explanations for their difficulties. "Do you have any idea what caused these feelings, what is actually behind these symptoms?" (p. 701) It is not clear whether Wolberg actually expects his clients to give reasonable explanations for their difficulties or whether he is simply orienting them to search for and to be alert to the causal factors. His own comment to the remark just quoted is: "Focusing her attention on the fact that there are causes that underlie the symptoms." (p. 701)

There are passages in this book that should be of particular interest for educators. For instance, on page 128 and in the following three pages there is a discussion of the teacher in psychotherapy. This treatment is by no means dogmatic and while Wolberg believes that the teacher is not equipped to do psychotherapy in the formal sense, teachers can profit a great deal in understanding some of the principles of psychotherapy in dealing with emotional problems that their pupils present.

The chapter on supervision in the psychotherapeutic process has immediate application to the problem of supervision in education and the dynamic principles of supervision which Wolberg enunciates applies to supervision in education as much as to supervision in psychotherapy. On page 650 Wolberg gives a list of fourteen propositions involved in the process of learning of psychotherapy, and the student of learning should gain from these valuable suggestions about some of the dynamics of the learning process. In passing, it may be added that Wolberg gives a very fair and well-informed discussion of the problems relating to the practice of psychotherapy by clinical psychologists. Percival M. Symonds

Teachers College, Columbia University

George R. Klare and Byron Buck. *Know Your Reader*. New York: Hermitage House, 1954, pp. 192. $2.95

Presumably, writing is done to communicate with a reader. Unless a writer knows the reader and his interests, he may well end up by talking to himself or to nobody. But if he knows his reader, he is able to write with some consideration for him. The authors of this treatise state that they have concentrated on kinds of information that can help a writer to meet his reader. Much of the material discussed is derived from application of readability formulas. Advantages as well as limitations of the formulas are noted. In the introductory material, consideration is given to the wide range of application of formulas along with some results.

The next series of chapters deal with the background for, and history of, readability research. Next comes a discussion of the reader himself. Without this, readability formulas are of little use. A series of three chapters deal with the technical background necessary for understanding formulas and their place in a writer's equipment. This is followed with discussions of preparation for writing and some aspects of style that make writing readable. Too little attention is given to format in relation to readability. Note is made of pitfalls which can arise from misunderstanding of formulas and their misuse. It is emphasized that even when formulas are employed properly, they leave unmeasured many of the important aspects of readable writing. A final chapter discusses the trend of contemporary reading research. Much of the material presented here is concerned with viewpoints and theory rather than research.

Certain statements in text as well as examination of the bibliography suggest that the authors have failed to note some of the recent experimental findings on scientific typography. Thus, (p.

132) the statement that all-capitals and italics tamper very little with set reading habits; (p. 134), that italics are no harder to read than ordinary Roman type; and (p. 133), that differences in perceptual span are relatively unimportant, as are differences in interfixation speed of movement.

The authors have presented a well-organized, well-written treatise. Their emphasis upon knowing the reader if communication through writing is to be effectively achieved is basic. Their evaluation of readability formulas and of the use and limitations of such formulas is a contribution of high value. Writers of textbooks, trade journals, materials for communications to workers in industry, and other books and magazine articles should profit by careful attention to the contents of this book. MILES A. TINKER

University of Minnesota

C. C. ROSS. *Measurement in Today's Schools.* Revised by Julian C. Stanley. (Third Edition). New York: Prentice-Hall, Inc., 1954, pp. 485.

The Ross text first appeared in 1941, and was revised in 1947. Following Professor Ross' death, Dr. Stanley prepared this third edition. No major changes in the pattern of the book or the point of view have been made since the first edition. The largest number of chapters bear on the use of test results, with two chapters on test construction, one on statistics, and one on characteristics of tests. In contrast to many other texts in this field, no space is given to systematic coverage of published tests or to problems associated with testing various content areas or types of behavior. Chapters on graph-making and public relations are offered.

Students have liked the Ross text for its definiteness. It gives specific procedures for item preparation, and covers many topics of interest to administrators and supervisors. On the other hand, the text seemed conservative even at its first appearance, and the revisions have preserved this character. The student is told many things which must seem to him purely academic in interest, and some chapters seem so to the reviewer also. The style is one of old-fashioned scholarship, with long and uninspired quotations, numerous tables, and footnotes by the dozen. On the positive side, Ross' treatment was sound, in the light of the knowledge of the 193[illegible] Stanley has clearly undertaken to change the book as little as possible. He has added new footnote references, has greatly modernized the chapter reading lists, and has added paragraphs at a few points. Several illustrative tables have been replaced with better examples. He has judiciously removed paragraphs by Ross, but has never rewritten material even where Ross' view is questionable. One change in the text increases an unfortunate overemphasis on item analysis as a way of determining test content. The interpretation of correlation is damaged by introducing a chart in which the Fisher z is taken as showing "the amount of relationship represented" by a given r, and by omitting Ross' reference to Taylor and Russell. To compensate for the absence of new ideas in the main chapters, Stanley has added one chapter on present trends. This material is too cursory to be significant; one simply cannot deal intelligently with factor analysis in five sentences, no matter how well chosen. This chapter does provide reassurance, lacking elsewhere in the book, that the testing field has not stagnated in recent years.

The chief fault in Ross' book is one which could not be remedied by retouching. The book is oriented around the superficies of measurement, such as scoring of rearrangement items and using legible record forms, rather than around behavior and learning. This is made obvious by the fact that the objectives of instruction and their relation to testing are discussed only in one three-page section. Curricular validity is dismissed with the trivial and incredible comment that this idea has "shifted the center of gravity from the curriculum to the child."

Those instructors who have been satisfied with the Ross text will be grateful for a revision such as might have been expected from Ross himself. The revision will not, however, appeal to any new audience.

LEE J. CRONBACH

University of Illinois

THE JOURNAL OF

EDUCATIONAL PSYCHOLOGY

Volume 46 October, 1955 Number 6

THE RELATIONSHIP OF IMMEDIATE RETENTION OF TECHNICAL TRAINING MATERIAL TO CAREER PREFERENCES AND APTITUDES[1]

G. R. KLARE, L. M. GUSTAFSON, J. E. MABRY AND E. H. SHUFORD

University of Illinois[2]

This paper presents a study of the relationship of immediate retention test scores (for a technical training passage) to selected 'career preferences' (expressed interests) and to certain measured aptitudes of airmen in indoctrination training. The data were collected as one aspect of several experiments in the learning and retention of technical training materials, chiefly for the purpose of determining whether the variance of retention test scores attributable to expressed interest was likely to be sufficient to require control of the interest variable. Since previous experimentation had shown that aptitude was likely to be a very important contributor to variance in retention test scores, this relationship as well as that of measured aptitude to career preferences was also investigated.

That interest is of potential significance to the process of learning was recognized as early as the Seventeenth Century (*13*). The interests of the learner have even been made the cornerstone of certain later philosophies of education (*13*). Though it was not until World War I that psychologists and educators began to study interest systematically, since that time the relation of interests

---

[1] This research was supported in part by the United States Air Force under Contract No. AF 33(038)-25726, monitored by the Personnel and Training Research Center. Permission is granted for reproduction, translation, publication, use and disposal in whole and in part by or for the United States Government.

[2] Dr. Klare is now at Ohio University; Dr. Gustafson at Ok[illegible] A. and M. College; and Dr. Mabry at VA Hospital, Salt Lake City, Utah.

(including career preferences) to measured ability has been the subject of many articles. This literature is too vast for review here, but certain salient points relative to the study reported in this paper will be mentioned.

The relationship of interests and abilities has been studied by means of two rather distinct kinds of interest measures. The first ('expressed interests') is based on subjects' statements of their interests, the second ('inventoried interests') is based on the use of an interest inventory. The latter is a more refined measure both statistically and semantically. These two kinds of interest measures are in themselves not highly related; Anderson (*1*), Berdie (*2*), Brown (*3*), and DiMichael (*5*), among others, have found that the relationship ranges from low to moderate. In those studies in which correlation coefficients were computed, the values tend to be about +.50.

Both kinds of interest measures have been studied in relation to a variety of measures of ability and achievement. Because of this variety, a comparison of studies is difficult to make; both positive and negative relationships have been found for both kinds of interest indices, the correlations being low to moderate in degree (Strong, *12*). Typical recent examples of studies showing little or no relationship are those of Jackson (*6*), Klugman (*8*), and Townsend (*14*). The studies of Kopp and Tussing (*9*) and Laycock and Hutcheon (*10*) are typical of those showing moderate relationships (correlation coefficients up to about +.50). One of the highest correlation coefficients found was reported by Wesley, Corey and Stewart (*15*), who found that a measure of interest in physical science yielded a correlation of +.66 with success in engineering courses.

METHOD

The data reported here were collected at two separate times but in much the same way. The first study served as a pilot experiment; the second as a check on the results of the first, with a larger number of subjects. In both studies the subjects read a technical passage and then took a test covering the passage. Immediately afterward, the subjects were asked to indicate their preferences for a number of Air Force career areas. Mechanical and clerical aptitude scores were also available for these subjects.

Immediate retention tests.—The tests used were of the four-alternative, multiple-choice form and covered a 1206-word printed lesson on the induction and cooling systems of an aircraft engine. The material was rather highly technical in nature and was one of the series of lessons used in training Air Force mechanics at Chanute AFB.

In the first study, the standard version of this material was read by the subjects and two test forms, called PD and PE, were used. In the second study, the standard version was again used along with several additional versions that contained the same content but differed in manner of presentation of the material. A test called form G, made up of items selected from tests PD and PE, was used with all versions in the second study. Test scores on the versions used did not differ significantly.

Aptitude indices (stanine scores).—Air Force personnel records contain aptitude indices on a 'stanine scale' (from 1 to 9) for all airmen. These indices, based on the Airman Classification Test Battery (*4*), cover aptitudes in eight areas: mechanical, clerical, equipment operation, radio operation, technical specialty, service, crafts, and electronics. In both studies reported here only the mechanical and clerical indices were used in analysis, the former because the reading material was mechanical in nature, the latter because of its relationship to reading ability.

Career preference check list.—The career preference check list covered eight general career areas, with a short description of each, based on the eight aptitude areas covered by the Airman Classification Test Battery. These areas were described as common to civilian life, and subjects were told to indicate with a '1' their first preference, '2' their second preference, etc., down to '8' as the least preferred area. They were told to answer as if they had equal chances for training and experience—and equal chances for pay and advancement—in each area. As in the case of the aptitude data, only the mechanical and clerical preference data were used in analysis.

Subjects.—Subjects for both the first and second studies were airmen in basic training at Sampson Air Force Base, an indoctrination center. Both studies were carried out during the early part of the basic training program, just after the regular classification tests had been administered.

For the first study, there were one hundred fifty-nine subjects who took retention test form PD and one hundred fifty who took test form PE. For the second study, there were four hundred forty subjects (all took the form G test).

RESULTS

Mechanical career preference and retention test score for a mechanical training lesson.—First study: It will be recalled that the first study was a pilot study, with relatively small N's (one hundred fifty and one hundred fifty-nine) in two test groups (each taking a different test form). It was noted in the first study that the mechanical work area was by far the most preferred of all the career areas listed; there was, in consequence, a relatively large number of subjects who listed mechanical work as a high preference and a relatively small number who listed it as a middle or low preference. In order to get about equal N's for high and low preference groups, high preference airmen were defined as all subjects who gave mechanical work a preference of 1, and low preference airmen as all subjects who gave mechanical work a preference of 6, 7 or 8.

Examination of the retention test score data for the high and low preference airmen showed that the former had somewhat higher test scores than the latter (in the case of both the PD and PE test groups). However, *t*-tests showed that the mean differences were not significant.

Second study: A 3 × 8 disproportionate analysis of variance was computed on the immediate retention test (form G) scores of the four hundred forty subjects. Three levels of mechanical work preference (1–2, 3–5, 6–8) and eight levels of mechanical aptitude (1–2, 3, 4, 5, 6, 7, 8, 9) were used as the independent variables (see Table 1). Though inspection of the data showed that higher preference was related to somewhat higher immediate retention test scores just as in the first study, the variance attributable to mechanical work preference in the analysis was not significant at the .05 level. The variances attributable to mechanical aptitude and to interaction were significant beyond the .05 level, but no consistent pattern of interaction effect was apparent.

Clerical career preference and retention score for a mechanical training lesson.—Analysis of the data for clerical career preferences and retention scores showed there was little relationship between the two in either the first or the second study.

Mechanical aptitude and retention score for a mechanical training lesson.—Table 1 is consistent with other work not reported here in

TABLE 1.—ANALYSIS OF VARIANCE OF IMMEDIATE RETENTION SCORES OF 440 SAMPSON AIRMEN IN RELATION TO MECHANICAL WORK PREFERENCES AND MECHANICAL APTITUDE INDEX SCORES

Source of variation	Sum of squares	df	Mean square	F	P
Mechanical work preferences	15.76	2	7.88	.25	$>.05$
Mechanical stanine scores	21,879.20	7	3,125.60	99.82	$<.05$
Interaction	919.14	14	65.65	2.10	$<.05$
Within	13,023.61	416	31.31		
Total	35,837.71	439			

showing that mechanical aptitude index can be expected to account for significant retention test variance. To ascertain the degree of the aptitude-retention score relationship, several product-moment correlations were computed. For the first study, the correlation coefficients were +.74 for the PD (N = 150) and +.75 for the PE (N = 159) groups. For the second study, the four hundred forty subjects' scores were analyzed as four groups of one hundred ten each (all groups took the Form G test); the correlation coefficients were +.68, +.75, +.78, and +.83. The relationship between mechanical aptitude and retention test scores in these studies thus appears quite high and stable.

Clerical aptitude and retention score for a mechanical training lesson.—The relationship for clerical aptitude and retention test score, though lower than that for mechanical aptitude and test score, is also quite high and stable. For the first study, the correlation for the PD group was +.55 and for the PE group +.60. For the second study, the correlations were +.54, +.55, +.62, and +.67. Intercorrelations of mechanical and clerical aptitude indices for the two studies ranged from +.63 to +.83 (median value, +.78).

Mechanical career preference and mechanical aptitude.—First study: In the first (pilot) study, a 2 × 2 disproportionate analysis of variance[3] (high versus low preference × groups PD and PE) was computed on mechanical aptitude scores. It was found that the variance attributable to preference was significant beyond the .05 level; the variance attributable to replication (test groups) and that attributable to interaction were not significant. Examination of the relevant means showed that subjects with a high preference for mechanical work had lower aptitude scores than subjects with a low preference.

---

[3] The method of analysis was that suggested by Snedecor (*11*) for the case of disproportionate frequencies in subclasses. In addition to computing Bartlett's test for homogeneity of variance, the precaution suggested by Johnson (7) of computing chi-square to determine whether actual cell frequencies differ significantly from those expected was also followed in all disproportionate analyses.

Second study: In the second study, a 3 × 1 analysis of variance was computed on mechanical aptitude scores, using three levels of mechanical work preference (1–2, 3–5, 6–8). It was found that the variance attributable to mechanical work preference was again significant beyond the .05 level, as in the first study. The same trend of means was also found, the subjects with a higher mechanical work preference having a lower mechanical aptitude score than the subjects with a lower preference. Since the distribution appeared curvilinear, an eta coefficient was computed and found to yield a low value.

Clerical career preference and clerical aptitude.—First study: In order to get about equal N's for the high and low clerical preference groups, the high preference group was defined as all subjects with a preference of 1 or 2 and the low preference group as all subjects with a preference of 8 (clerical work was among the least preferred of the eight career areas). A 2 × 2 disproportionate analysis of variance (high versus low preference × groups PD and PE) was computed on clerical aptitude scores. It yielded variance attributable to preference significant beyond the .05 level; the variance attributable to replication (test groups) and that attributable to interaction were not significant. Examination of the means showed that subjects who had a high clerical preference had somewhat higher clerical aptitude scores than subjects who had a low preference. This relationship, it will be noted, is the opposite of that found for mechanical preference and mechanical aptitude.

Second study: A 3 × 1 analysis of variance was computed on clerical aptitude scores, using three levels of clerical work preference (1–3, 4–6, 7–8). The variance attributable to preference was found to be significant beyond the .05 level, which is in agreement with the results of the first study. Also, as in the first study, subjects with a higher clerical preference were found to have higher clerical aptitude scores than subjects with a low preference. An eta coefficient computed on these data yielded a low value.

DISCUSSION

These studies show some tendency, albeit non-significant, for higher mechanical preference to be related to higher retention test achievement in the mechanical area. The non-significant relationships in these particular studies may have been due to the observed high motivation of all subjects in the test situation (which may well not hold up through an entire course of training), to a test not quite sensitive enough to show clear differences, or to other reasons. Further work is needed to determine whether the relationships found here are of any general significance.

In contrast to the preference-retention test results, the aptitude-retention test results clearly indicate a relationship that is consistently high. While this is not surprising, it does suggest that in such studies where aptitude indices are available the variance attributable to them should be accounted for in statistical analysis. The magnitude of the relationship indicates that this variance should be considered even when experimental groups appear to differ only relatively slightly in aptitude.

Apart from these results, the studies indicate certain interesting relationships between work preferences and aptitude scores. While the relationships are low—both for mechanical work preference and mechanical aptitude and for clerical work preference and clerical aptitude—the two studies are in agreement in disclosing significant relationships. This is of particular interest since the relationship for the former pair of factors is negative, and for the latter is positive. While both negative and positive relationships have been found in the literature, the observation that most are low suggests that the direction of the relationships may in many cases have been due to chance. The consistency in the two studies reported here, however, suggests somewhat greater stability for the present results.

One possible reason for the low relationship between mechanical work preference and mechanical aptitude is the experience in, and attitude toward, mechanical work that younger males get in this 'machine-age' society. That is, mechanical work is looked upon more favorably by most persons than many service occupations, and most persons have at least some mechanical experience. The relatively high preference status of mechanical work compared to other work areas has already been pointed out. This high preference status seems to indicate that interest in mechanical work arises largely from factors other than aptitude (or ability) and is fairly well spread over the aptitude range. Thus a low correlation between interest and aptitude in this work area would be expected.

Consideration of preference and aptitude for other work areas suggests, further, why the mechanical preference-aptitude relationship may be negative. Those subjects in these studies who received high mechanical aptitude scores also generally received high aptitude scores in other areas (an example from these studies is the median correlation of +.78 between mechanical and clerical aptitude scores). It seems possible that usually only the more able subjects in other work areas have highest career preferences in those areas (the positive relationship between clerical aptitude and clerical work preference is an example). This would tend to remove many of the subjects with higher mechanical (as well as other) aptitudes and leave a relatively larger proportion of the lower aptitude subjects to choose mechanical work. The low mechanical preference-aptitude relationship would become negative to the extent that this occurs, since it would tend to leave the lower-aptitude subjects with the higher preferences for mechanical work.

SUMMARY AND CONCLUSIONS

The two studies reported here were designed to determine the relationships of score on an immediate retention test covering a mechanical training lesson, selected career preferences (expressed interests), and selected airman aptitude indices. Neither mechanical work preferences nor clerical work preferences showed any significant relationship to retention test scores on the lesson. There was a tendency for higher mechanical work preference to be related to higher retention scores, but the relationship was too small to reach statistical significance. Both mechanical and clerical aptitude indices, on the other hand, appear quite highly and consistently related to retention test scores (this is particularly true for mechanical aptitude).

In these studies, a low negative relationship was noted between preferences for mechanical work and mechanical aptitude score. The relationship between clerical preference and clerical aptitude was also low, but positive. It is interesting to note that two separate studies yield similar results, indicating that the relationships, while low, were nevertheless stable.

REFERENCES

1) Rose G. Anderson. "Subjective ranking versus score ranking of interest values." *Personnel Psychol.*, 1948, 1, 349–355.

2) R. F. Berdie. "Scores on the Strong Vocational Interest Blank and the Kuder Preference Record in relation to self-ratings." *J. appl. Psychol.*, 1950, 34, 42–49.

3) M. N. Brown. "Expressed and inventoried interests of veterans." *J. appl. Psychol.*, 1951, 35, 401–402.

4) J. T. Dailey. *Development of the Airman Classification Test Battery.* San Antonio, Tex.: Human Resources Research Center, Lackland Air Force Base, November 1948 (*Research Bulletin* 48-4).

5) S. G. DiMichael. "The professed and measured interests of vocational rehabilitation counselors." *Educ. psychol. Measmt.*, 1949, 9, 59–72.

6) J. A. Jackson. "A note on the crystallization of vocational interests." *J. soc. Psychol.*, 1947, 26, 125–130.

7) P. O. Johnson. *Statistical Methods in Research.* New York: Prentice-Hall, 1949.

8) S. F. Klugman. "The effect of schooling upon the relationship between clerical aptitude and interests." *J. genet. Psychol.*, 1945, 66, 255–258.

9) T. Kopp and L. Tussing. "The vocational choices of high school students as related to scores on vocational interest inventories." *Occupations*, 1947, 25, 334–339.

10) S. R. Laycock and N. B. Hutcheon. "A preliminary investigation into the problem of measuring engineering aptitude." *J. educ. Psychol.*, 1939, 30, 280–289.

11) G. W. Snedecor. *Statistical Methods.* Ames, Iowa: Iowa State Coll. Press, 1949.

12) E. K. Strong, Jr. *Vocational Interests of Men and Women.* Stanford Univer. Press, 1943.

13) D. E. Super and J. W. Dunlap. "Interest in work and play." In D. H. Fryer and E. R. Henry (Eds.), *Handbook of Applied Psychology.* New York: Rinehart, 1950.

14) A. Townsend. "Achievement and interest ratings for independent-school boys." *Educ. res. Bull.*, 1945, 43, 49–54.

15) S. M. Wesley, D. Q. Corey and Barbara M. Stewart. "The intraindividual relationship between interest and ability." *J. appl. Psychol.*, 1950, 34, 193–197.

PERSONALITY AND TEACHER EFFECTIVENESS: A PROBLEM IN THEORETICAL RESEARCH[1]

E. G. GUBA and J. W. GETZELS

The University of Chicago

The idea that teacher effectiveness can be identified and predicted by the application of the usual kinds of selection techniques developed by psychologists and others has resulted in a great amount of research. The results are inadequately represented even in such extensive bibliographies as those by Barr (*2*), and Domas and Tiedemann (*3*). Despite a large number of investigations, relatively little more is known now than was known in 1900. Many of the studies have failed to produce any significant results. Many others, especially those dealing with personality variables, have resulted in pedestrian findings, as for example, that pupils learn better from teachers who are kind, cheerful, and sympathetic to pupil problems. Still others, for which statistically significant findings may be claimed, tend to have little practical significance because they involve an unwieldy number of variables.

What is perhaps the most important reason for this failure to produce generally meaningful and useful results has been stated rather succinctly in the two recent reports (*8*, *9*) of the Committee on Criteria of Teacher Effectiveness of the American Educational Research Association, H. H. Remmers, Chairman. The committee's contention is simply that much past research has been conducted in a theoretical vacuum, seeking *ad hoc* solutions to immediate practical problems, with little regard for the long range usefulness of the findings. Moreover, many of the hypotheses which have been investigated have been based upon an oversimplification of the situation and have been tested by means of an inadequate methodology, leading to conclusions which do not make psychological or sociological sense. The committee concludes that only by predicating hypotheses on sound theory can one hope for useful, relevant, and widely applicable findings.

---

[1] This research was supported in whole or in part by the United States Air Force under Contract No. AF 18(600)-415 monitored by the Air Research and Development Command.

The application of theory to research in teacher effectiveness, or indeed, to any research, has, as the committee report implies, several distinct advantages. The initial tendency of the worker in a new field of knowledge is to observe. 'Get the facts' is a familiar exhortation. And this is not without justification, for while the researcher's ultimate purpose may be to explain the phenomena which he will observe, he must first know what these phenomena are. Unless this particular researcher is exceptionally sophisticated, however, he is likely to overlook a vital element in constructing his design for observation; namely, that it is humanly impossible to make objective observations of all elements in a phenomenal field; and, further, that it is unlikely that those phenomena which are observed will be accorded appropriate weights. Consciously or unconsciously, the observer makes a selection from among the available phenomena, and then ascribes differential weightings of relative importance to the various elements which he does select. But what criteria does he utilize in making this selection and ascription? Since such criteria will inevitably be applied, is the researcher not better off for having made them explicit before the task of observation begins?

But theory has a greater usefulness than merely to provide the setting in terms of which a research design may be formulated. A theory is more than taxonomic, it is dynamic; it attempts not only to describe but also to explain. And precisely because it contains such dynamic elements, theory leads directly to further hypotheses which may be tested in terms of their congruence with observational data. An adequate theory is thus self-extensive; i.e., as portions of the theory are tested and found to be valid or invalid, new hypotheses suggest themselves which require further investigation, and so on. The findings resulting from the tests of such hypotheses then serve not only the descriptive function but also, and more importantly, the explanatory function.

Finally, theory provides a basis for interpreting those observational data which are collected. Without the benefit of such a framework, two contradictory observations cannot be resolved; there is no *a priori* basis for supposing that one observation is more valid or relevant than the other. Without such a framework, a single observation which is 'significant' must be accepted at face value. Theory acts here in two ways: first, by flashing a red 'caution' signal when data are uncovered which do not make sense in terms of the theoretical formulation, and, then, by serving as arbiter in the attempt to resolve apparent inconsistencies. This is not to say that theory will always be taken as correct or that observed data which contradicts the theory as wrong. It simply means that when such inconsistencies occur, they will be apparent and further investigation will be indicated.

An opportunity to demonstrate rather sharply the differences between a theoretical approach and the more traditional procedures arose in connection with a broad study of teacher effectiveness carried out by the present authors in a military setting.[2] Although the primary purpose of the study was not to discover relationships between personality and teaching effectiveness, data which became available made it possible to investigate these relationships utilizing a rather large number of instruments. Of these instruments, six are pertinent to the present paper: Three tests of a type which have frequently been used in normative, non-theoretically oriented research, and three others for which it is possible to suggest theoretical links with effectiveness. The purpose of this paper is to describe the application of these six devices in a study of the effectiveness of Air Force officer-instructors, and to indicate the major points of difference in the findings resulting from the two approaches.

PROCEDURES

The setting of the study was the Air Command and Staff School, Air University, Maxwell AFB. This school consists of eleven teaching units, nine of which maintain courses while the remaining two simply serve as lecturer pools. The subject matter is both civilian and military in nature (e.g., law and logistics). Instruction is carried on by two distinct methods, lecturing and discussion, with most instructors doing both kinds of teaching. Two quite different kinds of teaching situations, which might be expected to make differential demands upon the personalities of the instructors involved, were thus available for study.

The sample.—The general sample consisted of two hundred and thirty-four officer-instructors. A varying proportion of this group (N = 156 to 215) was tested with the six instruments in question and, in addition, assisted in the development of effectiveness criteria. The sample approximated very well the total population of instructors both in terms of rank (which ranged from Chief Warrant Officer to Lieutenant Colonel) and assignment to the eleven divisions. A group of sixty-one superiors, each of whom had at least nominal charge of a section of instructors, was also utilized to assist in the development of effectiveness criteria.

---

[2] A preliminary report of the larger study may be found in (*5*).

The effectiveness criteria.—The effectiveness criteria consisted of various composites of ratings made of instructors by their fellow-instructors (colleagues) and by their superiors. This procedure had validity in the Air University situation for several reasons: (1) a rating system is familiar to military officers and is part of their daily routine, (2) lectures which an AC&SS instructor give must be rehearsed before an audience of fellow-instructors and/or superiors, and (3) mutual visitation of classes is encouraged and, indeed, almost required of AC&SS instructors. The raters thus had opportunity to observe the instructors whom they were rating and implicitly made such ratings almost every day.

Each instructor and superior was required to list in rank order of effectiveness as lecturers and/or discussion leaders (depending upon the kind of teaching done) all of the instructors who served in his division or section. Standard (T) scores were then computed for each instructor to reduce the data from separate courses to comparable terms, and from these scores three composites were derived as follows:

1) Effective and ineffective lecturers. Persons who achieved T scores in excess of 60 (effective) or less than 40 (ineffective) when rated as lecturers by either colleagues or superiors were retained, provided, that persons rated in excess of 60 by one group but less than 40 by the other were omitted.

2) Effective and ineffective discussion leaders. These persons were selected as above.

3) Effective and ineffective instructors in general. The two combined colleague and superior criteria, one for lecturers and one for discussion leaders, described in (1) and (2) above, were further combined by a similar standard scoring process to produce one general criterion.

Statistical tests made upon the resulting criteria indicate that there is a high degree of agreement between the two rating groups (superior and colleagues) as well as among the raters within a specific rating group in the naming of effective and ineffective instructors. Other tests indicate that the ratings are non-random and made according to a consistent criterion, whatever this may be.

The normatively oriented instruments.—Three well-known standard personality instruments of a type which have often been used in the traditional, normative research were included in the present study: The *Guilford-Martin Inventory of Factors GAMIN*, the *Guilford-Martin Inventory of Factors STDCR*, and the *Allport-Vernon-Lindzey Study of Values*. These instruments are not in any sense intrinsically normative and non-theoretical, of course, but they are often used in this fashion and are so used here. Whatever the value of these instruments may be at the theoretical level for other purposes, there is no reason to suppose that such a heterogeneous collection of variables could meaningfully be synthesized into a theory of teacher effectiveness.

Both Guilford-Martin instruments contain a relatively large number of items to be answered in terms of a Yes-?-No key. Ten scores are available:

1) G—general pressure for overt activity.

2) A—ascendency in social situations as opposed to submissiveness; leadership qualities.

3) M—masculinity of attitudes and interests as opposed to femininity.

4) I—lack of inferiority feelings; self-confidence.

5) N—lack of nervous tenseness and irritability.

6) S—social introversion-extraversion. Shyness, seclusiveness, tendency to withdraw in social situations vs. sociability.

7) T—thinking introversion-extraversion. An inclination to meditative or reflective thinking vs. extraversion orientation of thinking.

8) D—depression. Habitually gloomy and pessimistic mood vs. cheerfulness and optimism.

9) C—cycloid disposition. Strong, emotional fluctuations, tendency toward instability, vs. uniformity of moods, evenness of disposition.

10 R—rhathymia. A happy-go-lucky, carefree disposition vs. an inhibited, over-controlled disposition.

The *Study of Values* is based upon the six personality types originally proposed by Eduard Spranger, each of which assumes one of the following values as the basic motivational force:

1) Theoretical. The theoretical person is predominantly empirical, rational, and critical. He sees the systematization of knowledge as his highest goal.

2) Economic. The economic man is interested primarily in utility.

3) Aesthetic. The aesthetic man is interested in beauty, particularly as developed in form and harmony.

4) Social. The social man has for his highest value the love of people, in the altruistic or philanthropic sense. He is interested in human relationships.

5) Political. The political man sees his highest value in power. Competition, struggle, influence, and renown are characteristic of his motives.

6) Religious. This variable is not religion in the usual sense, but a kind of mysticism in which the individual seeks to relate himself to the universe as a whole. The essential value is unity.

All three standard instruments were scored in accordance with published directions (*1*, *6*, *7*).

The theoretically oriented instruments.—Unfortunately, there is as yet no comprehensive theory of teacher effectiveness which would permit a straight-forward development of hypotheses amenable to test. It is possible, however, to select instruments which, to borrow a term from the Remmers committee, may be 'psychologized', i.e., for which a reasonable connection with effectiveness may be postulated on the basis of relevant psychological theory. It is in this sense that the instruments to be described in this section are theoretically oriented.

Whatever the teacher may teach, it is obvious that the teaching is carried on in the context of an interpersonal setting. It is this factor which, more than any other, accounts for the crucial importance of teacher personality in mediating the teaching-learning process. The teacher cannot force the pupil to learn; what he can do is to produce a situation which the pupil will find conducive to learning. To relieve the teaching process of its affective elements is to reduce it to a sterile, highly intellectualized procedure which the pupil is unlikely to find encouraging.

The establishment of a genuinely good working relationship between teacher and pupil thus requires an interaction not only at the intellectual level but also at the level of personality. Consider then the teacher who is unable to enter into such a relationship. Any one of several modes of behavior may result. The teacher may respond to the everyday frustrations of the classroom with some kind of aggressive reaction. On the other hand, the reaction may be suppressed, leaving the teacher to act not in terms of his own motivations and values but in terms of some imagined standard of how a teacher 'should' act. The second reaction may be worse than the first, for while a hostile teacher may be unpleasant, at least his true feelings are evident; the latter teacher lacks integrity and hides his true feelings behind a misleading facade.

How can these two reactions, which may be labelled frustration-aggression and reaction-suppression, be measured? Rosenzweig has made an analysis of frustration and has developed an instrument, the *Rosenzweig P-F Study*, which is designed to measure six separate variables found to be associated with frustration behavior (*10*, *11*). When the test is scored[3] in terms of the standardized procedures outlined in the Rosenzweig manual (*12*), six measures associated with these six frustration variables are obtained as follows:

Direction of Aggression.

E—extrapunitiveness, a measure of the extent to which aggression is turned onto the environment.

I—intropunitiveness, a measure of the extent to which aggression is turned by the subject upon himself.

M—impunitiveness, a measure of the extent to which aggression is evaded in an attempt to smooth over the incident.

Type of Reaction.

OD—obstacle dominance, responses in which the barrier occasioning the frustration stands out.

ED—ego defensiveness, responses in which the ego of the subject predominates.

NP—need persistence, responses in which the solution of the problem is emphasized.

The second reaction, reaction-suppression, is not so easily measured. Many of the suppressed attitudes, values, and basic dispositions are, of course, not apparent even to the subject himself. Moreover, the subject is not likely to verbalize publicly even those suppressed attitudes which are conscious. Hence, measurement devices must be used which tap deeper levels than those readily available through conscious recall, and whose purpose is not immediately apparent to the subject.

A suitable technique for the purpose at hand is the method of paired direct and projective questionnaires, devised by one of the authors (*4*). Depending upon the instructions and the item form, it is possible through the use of two such paired instruments to elicit information which bears upon both the overt and the covert attitudes of the subject toward the same object, and hence makes possible the measurement of the disparity between manifest and latent feelings.

---

[3] The authors are indebted to Francis J. Van Bortel of the Committee on Human Development, University of Chicago, who scored the Rosenzweig tests.

Considerable evidence exists (*4, 13*) which shows that subjects respond to two such forms in quite different ways. Typically, the response to the projective instrument is drawn from the deeper layers of personality, and may, especially in the case of a negative response, represent an attitude or feeling which is suppressed under normal circumstances. Responses to the direct instrument may in part, at least, reflect the subject's beliefs about how he 'should' act. To the degree to which the responses to the two forms are similar, the subject is relatively spontaneous and adjusted; while to the extent to which they differ, the subject is relatively repressed and maladjusted.

On the basis of such reasoning, a number of specific hypotheses[4] may be drawn concerning probable relationships between the various test scores and teacher effectiveness:

The *Rosenzweig P-F Study*, Direction of Aggression:

1) Extrapunitiveness, which may be interpreted as the tendency to place blame upon the environment when frustrations occur, is linked to ineffectiveness.

2) Intropunitiveness, which may be interpreted as the tendency to accept responsibility for frustrations that occur, is linked to effectiveness.

3) Impunitiveness, which may be interpreted as the tendency to avoid placing blame altogether, is linked to effectiveness.

The *Rosenzweig P-F Study*, Type of Reaction:

4) Obstacle dominance, which may be interpreted as the tendency to focus attention upon the obstacle occasioning the frustration, is linked to ineffectiveness.

5) Ego defensiveness, which may be interpreted as the tendency to focus attention on self, is linked to ineffectiveness.

6) Need persistence, which may be interpreted as the tendency to focus attention on the solution to the problem, is linked to effectiveness.

---

[4] These hypotheses may break down in the case of a sample which deviates extremely from national norms for one or more of the test variables. Comparison of the present sample with national norms indicates that the vast majority of subjects, including those labelled as 'ineffective', are well within the limits of normality; indeed, most of the present subjects exceed the national norms in the direction of preferred behavior.

The *Paired Direct and Projective Questionnaires.*

7) A high negative score on the projective form is linked to ineffectiveness.

8) A high negative score on the direct form is linked to neither effectiveness nor ineffectiveness, since both groups may have the same stereotypic beliefs about desirable behavior for the teacher.

9) A large disparity between negative scores on the projective and direct instruments (P-D), indicating a lack of spontaneity and integrity, is linked to ineffectiveness.

FINDINGS

For each of the sixteen normatively oriented variables and the nine theoretically oriented variables, comparisons were made between effective and ineffective lecturers, discussion leaders, and instructors in general by means of Fisher's t test. The findings for the normative variables are given in Table 1, and for the theoretical variables in Table 2. Both tables indicate the group obtaining the higher mean score for each variable, together with the associated level of significance.

The difference between the two tables which is noted immediately is the relatively greater proportion of significant differences which were uncovered with the theoretical instruments. Of the sixteen normative variables of Table 1, only 1 or 6.3 per cent are significant for lecturers, 6 or 37.5 per cent for discussion leaders, and 2 or 12.5 per cent for instructors in general. On the other hand, of the nine theoretical variables of Table 2, 67.0 per cent are significant for all three groups—lecturers, discussion leaders, and instructors in general. If the criterion of good research were simply the efficient selection of instruments, that is, selection in such way that a maximal number of significant variables were uncovered in proportion to the total number of variables tested, research conducted with theoretically selected instruments would seem to have a marked advantage.

A more useful comparison of the data provided by the two classes of instruments may be made in terms of interpretability, however. Normative instruments are generally selected because of the broad character of the variables under test, while theoretical instruments are selected on the basis of hypotheses developed from the application of relevant psycho-sociological theory. The data of Table 2 are thus immediately interpretable; the only question is whether or not the data tend to confirm or disprove the hypotheses. The

TABLE 1—DIRECTION AND SIGNIFICANCE OF THE DIFFERENCE OF MEAN SCORES ACHIEVED ON SIXTEEN NORMATIVELY ORIENTED PERSONALITY VARIABLES BY EFFECTIVE AND INEFFECTIVE LECTURERS, DISCUSSION LEADERS, AND INSTRUCTORS IN GENERAL

Variable	Lecturers		Discussion Leaders		Instructors in General	
	Group with Higher Mean Score	Level of Significance	Group with Higher Mean Score	Level of Significance	Group with Higher Mean Score	Level of Significance-
The Guilford-Martin Inventory of Factors GAMIN						
General Activity	Effective	NS	Effective	NS	Effective	NS
Ascendance	Effective	NS	Effective	NS	Effective	NS
Masculinity	Effective	NS	Effective	.10	Effective	NS
Inferiority (freedom from)	Ineffective	NS	Effective	.10	Effective	NS
Nervousness (freedom from)	Effective	NS	Effective	NS	Effective	NS
The Guilford-Martin Inventory of Factors STDCR						
Social Intraversion	Ineffective	NS	Ineffective	NS	Ineffective	NS
Thinking Intraversion	Effective	NS	Ineffective	NS	Ineffective	NS
Depression	Effective	NS	Ineffective	.05	Ineffective	NS
Cycloid Disposition	Effective	NS	Ineffective	.01	Ineffective	.10
Rhathymia	Effective	.10	Ineffective	NS	Effective	NS
The Allport-Vernon-Lindzey Study of Values						
Theoretical	Effective	NS	Effective	NS	Effective	NS
Economic	Ineffective	NS	Ineffective	.05	Ineffective	NS
Political	Ineffective	NS	Effective	NS	Ineffective	NS
Aesthetic	Effective	NS	Effective	.05	Effective	.02
Social	Ineffective	NS	Effective	NS	Effective	NS
Religious	Ineffective	NS	Ineffective	NS	Ineffective	NS

TABLE 2—DIRECTION AND SIGNIFICANCE OF THE DIFFERENCE OF MEAN SCORES ACHIEVED ON EIGHT THEORETICALLY ORIENTED PERSONALITY VARIABLES BY EFFECTIVE AND INEFFECTIVE LECTURERS, DISCUSSION LEADERS, AND INSTRUCTORS IN GENERAL

Variable	Lecturers		Discussion Leaders		Instructors in General	
	Group with Higher Mean Score	Level of Significance	Group with Higher Mean Score	Level of Significance	Group with Higher Mean Score	Level of Significance
The Rosenzweig P-F Study						
Extrapunitiveness	Ineffective	.05	Ineffective	.10	Ineffective	.05
Intropunitiveness	Effective	.01	Effective	.02	Effective	.02
Impunitiveness	Effective	NS	Effective	NS	Effective	NS
Obstacle Dominance	Ineffective	.10	Ineffective	NS	Ineffective	NS
Ego Defensiveness	Ineffective	NS	Ineffective	.02	Ineffective	.02
Need Persistence	Effective	.05	Effective	.01	Effective	.01
The Sentence Completion Tests						
Projective Negative Score	Ineffective	.10	Ineffective	.01	Ineffective	.01
Direct Negative Score	Effective	NS	Ineffective	NS	Ineffective	NS
Difference Score (P − D)	Ineffective	.02	Ineffective	.01	Ineffective	.02

following conclusions may thus be drawn immediately:

Hypothesis 1: Extrapunitiveness is linked to ineffectiveness. Hypothesis confirmed at the .10 level for discussion leaders, at the .05 level for lecturers and instructors in general.

Hypothesis 2: Intropunitiveness is linked to effectiveness. Hypothesis confirmed at the .02 level or better for all three groups.

Hypothesis 3: Impunitiveness is linked to effectiveness. Hypothesis in abeyance. No significant differences found, although the observed differences are all in the postulated direction.

Hypothesis 4: Obstacle dominance is linked to ineffectiveness. The hypothesis is provisionally confirmed for lecturers at the .10 level, non-significant but in the postulated direction for the other two groups.

Hypothesis 5: Ego defensiveness is linked to ineffectiveness. Hypothesis verified at the .02 level for discussion leaders and instructors in general, non-significant but in the postulated direction for lecturers.

Hypothesis 6: Need persistence is linked to effectiveness. Hypothesis verified at the .05 level for lecturers, at the .01 level for the two other groups.

Hypothesis 7: A high negative projective score is linked to ineffectiveness. The hypothesis is verified at the .10 level for lecturers, at the .01 level for the other two groups.

Hypothesis 8: A high negative direct score is related to neither effectiveness nor ineffectiveness. The hypothesis is verified in that (a) there are no significant differences, and (b) the direction of the differences is reversed for lecturers and discussion leaders.

Hypothesis 9: The difference between projective and direct negative scores is linked to ineffectiveness. Hypothesis verified at the .01 level for discussion leaders, at the .02 level for the other two groups.

Thus, it is clear that there is supportive evidence for each of the *a priori* hypotheses, although this evidence is not always significant in the statistical sense. In this connection, an example of the usefulness of theory in interpreting empirical data may be found in the case of the variable 'impunitiveness'. Hypothesis had indicated that this should be linked with effectiveness, and indeed it was, although not at statistically significant levels. If this variable had been one of several in a normative battery, an appropriate conclusion would have been that there is no relation between impunitiveness and effectiveness. The fact that it was included on a theoretical basis, however, gives one pause before dismissing it as actually non-significant. Are not the trends, weak as they may be, in the postulated direction? Does not this fact lend credence to the original formulation and suggest that possibly with another better-selected sample, or with some more refined technique, a significant relationship might not be found?

On the other hand, consider the findings reported in Table 1. Interpretation cannot be made simply by comparing findings to hypotheses, for such hypotheses do not exist. The data must then be accepted at face value, and explained purely on the basis of *a posteriori* rationalization. Consider for example the variable 'masculinity'. Why should effective discussion leaders have a high masculinity score, while for lecturers and instructors in general this variable seems to make no difference? To be sure, it may be recalled that the group studied here consists of Air Force officers among whom masculinity may presumably be prized, and that such masculinity may more easily be asserted in the face-to-face discussion situation than in a passive lecturing situation. But this is pure presumption, for which the data provide no test. This may be contrasted with the procedure of making the presumptions beforehand, and then arranging the design so as to make possible a test of their validity.

One final note may be made in connection with Tables 1 and 2. In so far as the data of Table 1 are significant at all, the significant differences tend to occur in connection with discussion leaders rather than lecturers or instructors in general. Since discussion leading obviously depends upon personal interaction to a much greater extent than does lecturing, this finding is not especially surprising. For Table 2, however, variables significant for discussion leaders are also (with one exception) significant for lecturers. If it may be asserted that more sensitive techniques are required to detect the influence of personality in the lecturing than in the discussion situation, it would appear that theoretical instruments directed specifically at particular variables are more successful in detecting such influences than are the more normative instruments.

CONCLUSIONS

On the basis of the analysis made at the beginning of this paper, and as this is supported by the admittedly limited and selected data of Table 1 and Table 2, the following relative advantages of theoretical over fact-finding or normative research in exploring relationships of personality and teaching effectiveness, may be noted:

1) The development of a theoretical framework sharpens the research objectives and makes specific the criteria of observation which are to be applied.

2) Variables which cannot meaningfully be organized into a psycho-sociological frame of reference are eliminated from the research design before they become wasteful of the researcher's resources.

3) The probability is increased that the research will be more efficient in the sense that a greater proportion of the variables included in the design will prove to be significant.

4) The usually complex task of interpretation is greatly simplified, sometimes involving little more than comparing findings with hypotheses.

5) Variables which are found to be non-significant, if they are in the theoretically postulated direction, are not likely to be eliminated without additional tests. Failure to detect important differences because of failure in methodological technique is thereby minimized.

6) The development of consistent hypotheses provides an opportunity not only for their test but also for the test of the underlying theory. The value of each unit of research is thereby increased.

7) The testing of specific hypotheses rather than a large number of unrelated variables increases the probability that differences which require sensitive handling will be uncovered.

8) Finally, the development of a systematic theory which can be tested at a large number of points will make possible the comparison of data derived from individual studies and will thereby expedite the emergence of psychologically and sociologically sound formulations.

REFERENCES

1) G. W. Allport, P. E. Vernon, and G. Lindzey. *Manual of Directions for a Study of Values*, Revised edition. Boston: Houghton-Mifflin Co., 1951.

2) A. S. Barr. "The measurement and prediction of teaching efficiency: a summary of investigations." *J. Experi. Educ.*, 1948, 16, 202–283.

3) S. J. Domas and D. V. Tiedemann. "Teacher competence—an annotated bibliography." *J. Experi. Educ.*, 1950, 19, 101–218.

4) J. W. Getzels. "The relationship between covert and overt levels of

attitude and personality organization as revealed by the method of paired projective and direct questionnaires." *Amer. Psychol.*, 1952, 7, 318 (abs.)

5) J. W. Getzels and E. G. Guba. "Rôle, rôle conflict, and effectiveness: an empirical study." *Amer. Sociol. Rev.*, 1954, 19, 164–175.

6) J. P. Guilford. *An Inventory of Factors STDCR: Manual of Directions and Norms*, Revised edition. Beverly Hills, Calif.: Sheridan Supply Co., n.d.

7) J. P. Guilford and H. G. Martin. *An Inventory of Factors GAMIN: Manual of Directions and Norms*, First revision. Beverly Hills, Calif.: Sheridan Supply Co., n.d.

8) H. H. Remmers (Chm.) *et al.* "Report of the Committee on the Criteria of Teacher Effectiveness." *Rev. Educ. Res.*, 1952, 22, 238–263.

9) ———. "Second report of the Committee on the Criteria of Teacher Effectiveness." *J. Educ. Res.*, 1953, 46, 641–658.

10) S. Rosenzweig. "The picture-association method and its application in a study of reactions to frustration." *J. Pers.*, 1945, 14, 3–23.

11) S. Rosenzweig, H. J. Clarke, M. S. Garfield, and A. Lehndorff. "Scoring samples for the Rosenzweig Picture-Frustration Study." *J. of Psychol.*, 1946, 21, 45–72.

12) S. Rosenzweig, E. E. Fleming, and H. J. Clarke. "Revised scoring manual for the Rosenzweig Picture-Frustration Study." *J. of Psychol.*, 1947, 24, 165–208.

13) J. J. Walsh. *Socialization Studied by a Sentence Completion Technique.* Unpublished Ph.D. dissertation, University of Chicago, 1953.

THE RELATIVE STABILITY OF CLASSROOM SOCIAL STATUS WITH UNWEIGHTED AND WEIGHTED SOCIOMETRIC CHOICES

NORMAN E. GRONLUND

University of Illinois

The use of the sociometric test, to determine the relative social status of pupils in classroom groups, is finding widespread use among teachers and research workers alike. One of the persistent questions concerning its use, in this regard, is whether or not the choices received should be weighted during the scoring. There are two common practices. One, is to give each choice a value of one regardless of level of choice; the other is to assign arbitrary weights to the various choice levels. In the latter case, the first choice is generally given a value equal to the total number of choices used and each succeeding choice declines in value by one. For example, with three choices a first choice is counted three points, a second choice two points, and the last choice one point. This weighting of choices is based on the assumption that the social significance of sociometric choices declines as the level of choice decreases and the decline in significance is generally proportionate to the weights assigned to the various choice levels. The weights selected, however, are arbitrary and without the support of experimental evidence.

In light of the above contradictory practices regarding the scoring of the sociometric test, it seemed desirable to investigate the problem further. Ideally, this would involve a determination of the social significance of the various choice levels and a comparison of the results with the arbitrary weights commonly assigned. Since this approach would be extremely difficult, if not impossible, it was decided to answer the above question in terms of the stability of the sociometric results. Specifically, this study is an attempt to determine the relative stability of unweighted and weighted social status scores using a sociometric test of three and five choices, with one criterion. It was decided that less than three choices would not give a sufficient spread of scores for a study of their stability and more than five choices would be unnecessary, since it has been shown (*6*) that there is no increase in stability beyond that number. The use of one criterion is in harmony with the common practice of classroom teachers. Incidental questions arose during analysis of the data and are answered later in this paper.

Although the arbitrary weighting of sociometric choices, in the determination of classroom social status, is a fairly common practice, none of the numerous investigations (*1*, *2*, *3*, *9*) of the stability of sociometric results were concerned with weighted scores. However, both Moreno (*5*) and Northway (*7*) implied the possibility of improving the stability of sociometric results, through weighting, by pointing out that the constancy of sociometric choices declines as the choice level decreases. A review of the literature has revealed no attempt to determine the relative stability of classroom social status based on unweighted and weighted sociometric choices.

METHOD OF INVESTIGATION

The design of the study is relatively simple. A sociometric test, requesting each pupil to select *the five classmates with whom he most prefers to work*, was administered to all of the pupils in nine public-school classrooms[1] on two different occasions. The first test was administered in the middle of January and the second test four months later. There was a total of 216 pupils from the fourth, fifth, and sixth grades included in the study. Of these, 118 were boys and 98 were girls. The nine classes ranged in size from 19 to 30 pupils.

ANALYSIS OF DATA AND RESULTS

The social status of the pupils in each class was determined by four different scoring methods for each administration of the sociometric test. These scoring methods are as follows:

5 choices (unweighted)—1 point for each choice
5 choices (weighted) —5 points for 1st choice
4 points for 2nd choice
3 points for 3rd choice
2 points for 4th choice
1 point for 5th choice
3 choices (unweighted)—1 point for each choice
3 choices (weighted) —3 points for 1st choice
2 points for 2nd choice
1 point for 3rd choice

---

[1] Grateful acknowledgment is extended to Miss Margaret Henderson, Director of Elementary Education, and the principals and teachers of the Champaign Public Schools for making this study possible.

TABLE I—CORRELATION COEFFICIENTS REPRESENTING THE TEST-RETEST STABILITY OF UNWEIGHTED AND WEIGHTED SOCIAL STATUS SCORES OVER A FOUR-MONTH PERIOD (RHO'S)

Classrooms		5 Choices		3 Choices	
Grade level	N	Unweighted score	Weighted score	Unweighted score	Weighted score
6	28	.85	.83	.82	.81
5	18	.83	.76	.70	.68
5	24	.78	.75	.68	.68
4	30	.74	.75	.72	.71
4	29	.74	.71	.62	.61
4	21	.73	.81	.72	.76
4	20	.73	.80	.69	.83
4	19	.72	.61	.48	.70
4	27	.70	.74	.61	.64
Mean		.76	.75	.67	.71

When the social status of the pupils had been determined for each administration of the test, the relative stability of the results obtained by the four scoring methods, over a four-month period, was determined by use of the Spearman rank-difference method of correlation. The results of this analysis are presented in Table I.

It will be noted in this table that there is no difference in the relative stability of unweighted and weighted choices at either the five-choice or three-choice level. However, inspection of Table I will reveal that the unweighted social status scores with five choices are consistently more stable than the unweighted social status scores with three choices. The difference between means of these two groups was found to be significant at the .01 per cent level. Apparently the social status of pupils is most stable where the results are based on five choices and the above arbitrary weighting of choices does not affect this stability. Since the stability of unweighted and weighted choices is practically identical, it was decided to determine the extent to which the two types of scores were comparable in indicating the social status of pupils. This was done by calculating rank-difference correlation coefficients between the social status of pupils, on the same test, using unweighted and weighted scores. These coefficients ranged from .90 to .97 with a median of .95, where five choices were used, indicating that the relative social status of pupils is essentially the same with both

TABLE II—STABILITY OF VARIOUS SOCIOMETRIC CHOICE LEVELS OVER A FOUR-MONTH PERIOD IN NINE LATER ELEMENTARY CLASSROOMS

Sociometric Choice Level	Per Cent	
	Md	Range
First	72	60 to 86
Second	59	50 to 71
Third	52	38 to 64
Fourth	45	26 to 63
Fifth	38	21 to 47

types of scores. Thus, there appears to be no advantage in spending the additional time required in the calculation of weighted social status scores, where the above weights are assigned.

Stability of Various Sociometric Choice Levels.—In the hope of developing a more refined weighting system that would improve the stability of social status scores, it was decided to analyze the relative stability of the various sociometric choice levels. This involved a comparison of the extent to which the various choices given on the first sociometric test were repeated on the second test, four months later. For example, it was noted on the first test, in each class, which classmembers received the *first choice* given by each pupil. The percentage of these classmembers that were chosen by the same pupils on the second test was then calculated and used as an index of stability for the *first choice*. The same procedure was used to calculate the stability of the other choice levels. Table II summarizes the results of the above analysis. It can readily be seen in this table that there is a declining stability from the first to the fifth choice. The stability of the first choice ranges from 60 per cent, in the class having the lowest stability, to 86 per cent, in the class having the highest stability, with a median of 72 per cent. In contrast, the stability of the fifth choice ranges from 21 per cent to 47 per cent with a median of 38 per cent. From the first to the fifth choice there is a steady decrease in median stability, although there is some overlapping in the range of stability scores at each choice level. Using the relative stability of the various sociometric choice levels as a guide, the following weighting systems were applied to the social status scores and the stability, over a four-month period, was determined for each system of weights.

Weighting System	First Choice	Second Choice	Third Choice	Fourth Choice	Fifth Choice
A	7	6	5	5	4
B	7	6	5	4	3
C	5	4	3	3	2
D	4	3	2	2	1
E	3	3	2	2	1
F	3	2	2	1	1
G	3	2	1	1	1
H	2	1	1	1	1

Of the above weighting systems, the 3–2–1–1–1 weights provided the greatest stability of social status scores. However, the rank-difference correlation coefficients ranged from .70 to .85 with a mean of .77, which is almost identical to the stability coefficients obtained with unweighted choices. Thus, it appears safe to conclude that the stability of sociometric scores, used to indicate the social status of pupils in the classroom, is not improved by the assignment of these arbitrary weights to the various choice levels.

Stability of High and Low Social Status Scores Using Five Unweighted Choices.—Although teachers and research workers are interested in the relative social status of all members of a given group, the extreme social status positions are of special importance. The development and utilization of the leadership potentialities of those pupils with high social status and attempts to improve the social acceptability of those pupils with extremely low social status is of central concern to teachers. Likewise, research workers investigating the correlates of social status direct most of their attention to the positions of leadership and isolation. Consequently, an analysis of the stability of high and low social status scores has practical significance for both groups.

In order to make the high and low social status groups comparable from class to class, Bronfenbrenner's (*3*) fixed frame of reference was applied. This procedure sets nine choices as the lower limit of the high chosen group (stars) and one choice as the upper limit of the low chosen group (neglectees) where five unweighted choices are allotted. On the basis of chance expectancy, these limits are significant at the .02 level.

To determine the stability of the high and low social status groups over the four-month period, the following formula was applied to these categories in each classroom group.

$$\text{Per Cent Agreement } \frac{2 \times \text{Number of Agreements}}{\text{No. in category (Test I)} + \text{No. in category (Test II)}}$$

The four-month stability of the high status group ranged from 44 to 100 per cent with a median of 60 per cent. Since it appeared that many of the pupils in the high chosen group were at the lower limit on one administration of the sociometric test and the loss of one choice would put them out of that category, it was decided to calculate the stability of the high status group with an allowable deviation of one choice below the lower limit. This analysis increased the stability to a range of 67 to 100 per cent with a median stability of 80 per cent.

Applying the above formula to the low status group, the four-month stability ranged from 25 to 100 per cent with a median of 67 per cent. Allowing a one choice deviation, similar to that for the high status group, increased the stability to a range of 57 to 100 per cent with a median stability of 80 per cent. Thus, the stability of the high social status group and the low social status group is comparable, with an average of 80 per cent of the pupils remaining in the same category, or one choice removed from it, over a four-month period. This high degree of stability, compared to the stability coefficients reported in Table I, indicates that the extreme social status scores are more stable than those in the center of the distribution. It also indicates that large fluctuations in social status are relatively rare where five unweighted choices are used.

Stability of Cross-Sex Choices Using Five Unweighted Choices.—Although it has been shown that there is a fairly high degree of stability of social status scores over a four-month period, where five unweighted choices are used, it is suspected that these stability coefficients are reduced, somewhat, by the instability of cross-sex choices. It is common knowledge that there is a cleavage between the sexes during the later elementary period, which in all probability would weaken any sociometric bonds existing between boys and girls. The validity of the above assumption was determined by separately analyzing and comparing the cross-sex choices of boys and girls on the two administrations of the sociometric test. Using the same formula used to determine the stability of the high and low social status groups, it was found that the constancy of boys' choices of girls, in the nine classes, ranged from 10 to 70 per cent with a median stability of 24 per cent. The stability of girls' choices of boys was even lower, with a range of 0 to 67 per cent and a median stability of 14 per cent. There can be little doubt that the instability of these cross-sex choices reduced the four-month stability coefficients, reported earlier, since approximately one-fifth of the choices allotted in the nine classes were given to members of the opposite sex. Specifically, the median per cent of cross-sex choices on the first administration of the test was 19 for boys and 22 for girls. On the second administration of the test the medians were 17 and 20 per cent, respectively.

DISCUSSION OF RESULTS

An analysis of the above data has revealed that the stability of pupils' social status, over a four-month period, is not improved by the arbitrary weighting of sociometric choices. It is recognized that there may be other reasons for weighting sociometric choices, but until experimental evidence can justify the value of assigning weights, it is suggested that each choice be given a value of one regardless of level of choice.

Since five choices have provided the most stable index of social status, with a single sociometric criterion, it appears advisable to use at least this number of choices where the measurement of social status is being considered. In his original use of the sociometric test Moreno (*5*) allotted two choices for the selection of seating companions. Since that time, choices have ranged from one to an unlimited number with little, or no, justification given for the number selected. Northway (*7*) has used and recommended three choices, but admits there is no experimental evidence to support this number. With the lack of agreement on the number of choices to be used, the feeling has arisen that the number of choices should be adapted to the choosing situation. For example, if committees of four are being formed, three choices is the logical number to allot to each pupil. This practice is based on the assumption that the choosing situation will be more realistic if the number of choices does not exceed the number of available positions in the situation under consideration. However, the reality of the choosing situation can be maintained with five choices by careful phrasing of the directions. In the above illustration, for example, the pupils could be told that five choices are needed since it is usually impossible to give all pupils their first three choices. This is in no way an attempt to mislead the pupils, but rather the solution to a very real problem in sociometric grouping. Besides providing a more stable index of social status, it will make it easier to organize groups where conflicts arise among the first three choices.

The fairly high degree of stability of social status over a four-month period, indicated in this study, must be interpreted in terms of the grade level and criterion used. Research has shown (*3*, *4*) that the stability of sociometric results, at the elementary level, increases from the first to the sixth grade. Consequently, social status at the early elementary level on the same criterion, would not be expected to attain the four-month stability indicated in this study. On the other hand, stability coefficients calculated solely from fifth- or sixth-grade populations would probably exceed those reported here. Selection of a different criterion would also influence the stability of social status to a certain degree. There is some evidence (*8*) that social status calculated from general criteria, such as the one used in this study, is more stable than that obtained with criteria based on specific skills. It appears that choices on a general criterion may be elicited by less modifiable characteristics related to the social attraction of the total personality, whereas choices on a specific criterion are determined more by situational factors that are constantly changing.

SUMMARY AND FINDINGS

This study was an attempt to determine the relative stability of unweighted and weighted social status scores using a sociometric test of three and five choices. This test, requesting each pupil to select *the five classmates with whom he most prefers to work*, was administered to all of the pupils in nine public-school classrooms on two different occasions, four months apart. An analysis and comparison of the data revealed the following findings.

1) There is *no* difference in the relative stability of social status based on unweighted and weighted sociometric choices, over a four-month period. This held true for both the three- and five-choice levels, using the arbitrary weights most commonly assigned. Rank-difference correlation coefficients between unweighted and weighted scores, on the same administration of the test, ranged from .90 to .97 with a median of .95, indicating that the relative social status of pupils is essentially the same with both types of scores.

2) The stability of the various choice levels on the sociometric test indicated that the first choice was most stable, with a steady

decrease in stability to the fifth choice. The median stability results were 72 per cent for the first choice; 59 per cent for the second choice; 52 per cent for the third choice; 45 per cent for the fourth choice; and 38 per cent for the fifth choice. Several weighting systems based on the relative stability of the various choice levels produced no increase, above that of unweighted choices, in the stability of social status over a four-month period.

3) The stability of social status based on five unweighted choices was consistently higher than the stability of social status based on three unweighted choices. The mean rank difference correlation coefficients for the two groups were .76 and .67, respectively. This difference between means was significant at the .01 per cent level.

4) The high and low social status scores, using five unweighted choices, were very stable over a four-month period. Using Bronfenbrenner's fixed frame of reference, it was found that of those pupils in the high and low social status categories on the first test, 80 per cent remained in the same category, or one choice removed from it, four months later.

5) The stability of cross-sex choices, over a four-month period, varied from class to class, but in general was relatively low. The stability of boys' choices of girls ranged from 10 to 70 per cent with a median of 24 per cent. The girls' choices of boys were even less stable with a range of 0 to 67 per cent and a median of 14 per cent. Since approximately one-fifth of the choices allotted in the nine classes were cross-sex choices, it was indicated that these choices contributed disproportionately to a reduction in the stability of the social status scores over a four-month period.

A discussion of the results leads to the suggestion that five unweighted choices be allotted to each sociometric criterion where the classroom social status of pupils is being determined.

REFERENCES

1) Merl E. Bonney. "The Relative Stability of Social, Intellectual, and Academic Status in Grades II to IV, and the Inter-Relationships Between These Various Forms of Growth." *Journal of Educational Psychology*, 34 (February, 1943), 88–102.

2) Merl E. Bonney. "The Constancy of Sociometric Scores and Their Relationship to Teacher Judgments of Social Success, and to Personality Self-Ratings." *Sociometry*, 6 (November, 1943), 409–424.

3.) Urie Bronfenbrenner. *The Measurement of Sociometric Status, Struc-*

ture and Development. Sociometry Monographs, No. 6. New York: Beacon House, 1945. Pp. 80.

4) John E. Horrocks, and Mae E. Buker. "A Study of the Friendship Fluctuations of Preadolescents." *Journal of Genetic Psychology*, 78 (June, 1951), 131–144.

5) J. L. Moreno. *Who Shall Survive?* Washington, D. C.: Nervous and Mental Disease Publishing Co., 1934. Pp. 440.

6) W. I. Newstetter, M. J. Feldstein, and T. M. Newcomb. *Group Adjustment: A Study in Experimental Sociology*. Cleveland: Western Reserve University, 1938. Pp. 154.

7) Mary L. Northway. *A Primer of Sociometry*. Toronto, Canada: University of Toronto Press, 1952. Pp. 48.

8) Mary L. Northway. *Appraisal of the Social Development of Children at a Summer Camp*. Toronto, Canada: University of Toronto Studies, Psychology Series, Vol. V, 1940. Pp. 62.

9) S. L. Witryol, and G. G. Thompson. "An Experimental Comparison of the Stability of Social Acceptability Scores Obtained with the Partial-rank-order and Paired-Comparison Scales." *Journal of Educational Psychology*, 44 (January, 1953), 20–30.

SEX DIFFERENCES IN ADJUSTMENT TO COLLEGE

JOHN SUMMERSKILL and C. DOUGLAS DARLING

Cornell University

A series of studies is being conducted at Cornell University on circumstances associated with good and poor adjustment in college students. These studies are focused on the characteristics of entering freshmen as related to subsequent performance in college. The subjects are freshmen who entered the University in September, 1948. McCormick (*12*) and Thompson (*19*) have reported on academic success and failure among students of arts and science in this class; Bentson (*1*) studied athletic success and failure among men in this class; and a report is in press (*18*) on the over-all incidence of medical disorders among these students. The present study concerns sex differences in performance demonstrated by the students of this class.

The extensive literature on psychological sex differences suggested that sex bears an important relationship to collegiate performance. Johnson and Terman have generalized from considerable evidence to sex differences in social attitudes and emotionality. Scheinfeld's "Women and Men" (*16*) describes, with experimental and anecdotal data, the biological and cultural determination of widespread sex differences. Freeman and Miles (*7*) review studies which demonstrate "significant differences in general interests, attitudes, pursuits, and occupations." A review by Schaffter and Woody (*15*) cites opinion and data on sex influences upon academic success and social status. Kinsey's second volume (*10*), based on a population which is predominately college educated, argues strongly for psychological differences associated with biological sexual expression. Recent large-scale surveys (*6*, *17*) have shown differences in the attitudes of male and female students towards college as an institution; and one comprehensive survey (*8*) found that such attitudinal differences are present and important beyond the undergraduate years.

It is clear that men and women display differences in percepts, thoughts, and feelings, when queried about themselves, other persons, or things. Sex differences in the organization of western society are clearly reflected in statistics on jobs and wages, outstanding artists, inventors, etc. Less clearly established are the nature and extent of differences in day-to-day behavior of men and women, e.g., success and failure in daily work, illnesses and injuries, use of leisure hours, etc. The writers believe that conclusions about the relative adjustments of men and women will be clarified by continuing research on performance variables of this kind.

METHOD

This study is based on the individual records of the entire class of 1,818 freshmen who entered Cornell University in September, 1948, for the regular four- or five-year undergraduate program. The data for the study included all available records pertaining to the college careers of these students through the succeeding four years, i.e., until June, 1952. These data were obtained from the following sources: office of the University registrar, offices of the various colleges at Cornell, Student Medical Clinic, University Athletic Association, Student Conduct Committee, students' yearbook for 1952. All data were coded and punched on separate IBM cards for each student. Precautions were taken so that confidential data would be accessible only to the investigators.

Cards were sorted into various performance categories as described below. Numbers of men and women in performance groups were tabulated, compared, and evaluated statistically by chi square tests.

RESULTS

There were 402 women among the 1,818 freshmen who matriculated in 1948, ie., 78% of the class was male, 22% female.

Attrition.—Findings for the class as a whole show that 727 out of 1,818 students had left the University without graduating by the spring of 1952. That is, the net attrition rate was 40%. Among those who left, there were clear-cut scholastic reasons ("busted out") in about 25% of the cases; and no apparent scholastic reasons existed (no "probation" or "warning") in about 50% of the cases. Scholastic factors may or may not have been important in the remaining cases who withdrew after having been on academic probation.

There were no differences between male and female students in rate of attrition. Women accounted for 22% of the class in 1948 and for 21% of the class in 1952. However, only 11% of the 175 outright failures were women, whereas women comprised 38% of the group of 386 students who left without a history of scholastic difficulty. The difference between the two groups represented by the latter percentages is statistically significant (*p* value from chi square less than .01).

Academic Achievement.—In the class there were 136 students who had been placed on academic probation in two or more different semesters during their four years at the University. This group of inferior scholars was compared with a group of superior scholars who stood in the top tenth of their respective schools according to cumulative grade averages at the end of the junior year. Academic achievement was assessed in this manner to avoid the influence of different grading systems found in different colleges of the University.

The percentage of women among inferior scholars was found to be disproportionately low (5%) and the percentage of women among superior scholars was found to be disproportionately high (30%). These figures represent a difference which is significant at the .01 level of confidence.

Campus Activities.—This class contributed 84 leaders (president or chief officer) to the major campus organizations, fraternities, and sororities. The student yearbook revealed another group of 89 students who did not report participation in a single club, society, committee, or athletic team, during four years at the University.

Women were represented in 20% of the leadership posts, i.e., were represented according to population. However, only 8% of the non-participants were women. The latter represents a significant deviation from the number of women in the class (*p* value from chi square less than .05).

Health.—The health data showed that 95% of this class visited the Student Medical Clinic during their college careers. On the average, each student experienced six different medical disorders (illness or injury) while at college. There were 126 students with no more than two medical complaints during the full four-year period. There were 112 students with at least 15 complaints during this same period.

The group which had relatively good health records was 17% female. The group with the relatively poor health records was

29% female. This sex difference is statistically significant at the .05 level of confidence. The data indicate that the difference holds for emotional as well as physical illness although numbers in the former category were too small for statistical evaluation of differences.

DISCUSSION

Significant differences in the performance of men and women were observed in each of the areas investigated, ie., academic progress, participation in campus activities, and health status. This indicates that there is a generalized sex difference in adjustment to college, but the direction of the difference depends upon the specific area under consideration. Thus, it is the women who tend more to poor health records, but the men who tend more to poor scholastic records and to non-participation in campus activities. Disrupted performance is found in both men and women; the differences are in the kinds of performance disrupted. This fact is noteworthy because standardized tests have consistently led to the generalization that women are less emotionally stable than men (*9*).

The differences reported here are statistically significant but not large. It is likely that sex accounts for only a minor part of the variation between individuals in the areas studied. In the present-day college environment sex appears to modify rather than critically determine behavior in these areas.

With regard to academic progress, our findings are in agreement with those reported by McNeeley (*13*). He found that fewer women leave college due to academic failure and more leave for "intangible causes." Among these were included "such terms as 'unable to concentrate on work', 'lack of interest', and 'inability to coordinate efforts'." These statements imply a sex difference in motivation for completing college which is also reflected in questionnaire data obtained at Cornell (*17*) and elsewhere (*6*). These surveys showed that more men undergraduates are vocationally oriented in their reasons for attending college. It is likely that fewer women assign "college degree" the priority assigned by men in striving for personal security. Some women undoubtedly withdraw from college simply because other avenues to personal security gain in priority and one cannot assume that withdrawal represents failure or regress in personal adjustment.

Marriage is frequently mentioned as a goal which replaces "college degree" and consequently accounts for the withdrawal of many women from college. The evidence on this is still equivocal. Few women explicitly say that they go to college primarily to further marriage plans (*17*). Nor have studies on attrition detected a substantial number of women explicitly leaving college to be married (*4*, *5*, *13*, *20*). Further, the number of unmarried women among college graduates is known to be disproportionately high (*8*), but we do not know whether this holds for those who leave before graduating. Further research on this topic is needed.

With regard to academic achievement, our findings on failures, probations, and high standing, all favor women students at Cornell. The fact that the number of men admitted to this class was four times the number of women raises the definite possibility that the women freshmen were, on the average a more select group of scholars. In any event, earlier prejudices about women's inability to handle college work are flatly contradicted by these and similar reports (*6*, *15*).

The finding that comparatively few women fail to participate in campus activities may also be partly explained by the local situation. All women at Cornell live in dormitories or sororities with comprehensive recreational and social programs. A substantial number of men in this class lived off-campus after their freshman year, and many men had part time employment off-campus. Also, veterans (15%) and married students (3%) were predominately men and their recreational and social activities were, in all likelihood, less closely identified with the University. The central fact, however, is that co-education clearly extends to organized student activities on the present-day campus.

The obtained sex differences in health status cannot be interpreted with finality. A higher morbidity (and lower mortality) rate among women has been reported for non-college populations (*2*, *3*). Whether this represents a true difference in the incidence of somatic disease (*16*) or reflects a difference in motivation (*11*) is not known. It is possible that more women visit the Medical Clinic at Cornell because all women live in University buildings with a well-knit counseling system which places emphasis on preventive health measures. However, evidence from a recent study of mononucleosis (*14*), a disease which is relatively prevalent and troublesome in the college-age group, indicates that the incidence is truly higher among women and that physical signs accompany the more frequent subjective complaints of women students. In any event, there is little doubt that the students who averaged more than two medical complaints each semester throughout their college careers were unduly handicapped in their day-to-day existence. A study focused on the men and women in this category would shed light on psychodynamics correlated with this sex difference.

SUMMARY

Records pertaining to the college careers of 1,818 students in the class entering Cornell University in 1948 were analyzed. Seventy-eight per cent of the class was male, 22 per cent female. When the two groups were compared, the percentages of women among outright scholastic failures, inferior scholars, and non-participants in campus activities, were found to be disproportionately low. The percentages of women among non-academic withdrawals from the University, superior scholars, and students in poor health, were found to be disproportionately high. Findings are interpreted in the light of previous research on sex differences in collegiate performance.

REFERENCES

1) Thomas G. Bentson and John Summerskill. "The Relation of Personal Success in Intercollegiate Athletics to Certain Aspects of Personal Adjustment," *Res. Quarterly*, In press.

2) Antonio Ciocco. "Sex Differences in Morbidity and Mortality," *Quart. Rev. Biol.*, 1940, 15, 59–73, 190–210.

3) S. D. Collins. "Cases and Days of Illness Among Males and Females," *U. S. Publ. Health Serv. Report*, 1940, 55, 47–93.

4) Jean Failing. *A study on attrition among freshmen who entered the N. Y. State College of Home Economics, 1938–1947*. Mimeographed report, Cornell University, 1951.

5) Daniel D. Feder. "Student Population," in *Encyclopedia of educational research*, Walter S. Monroe (Ed.), New York: The Macmillan Co., 1950.

6) Norman Frederiksen and W. B. Schrader. *Adjustment to College*, Princeton: Educ. Testing Service, 1951.

7) Frank S. Freeman and Catherine C. Miles. "Sex Differences," in *Encyclopedia of educational research*, Walter S. Monroe (Ed.), New York: The MacMillan Co., 1950.

8) Ernest Havemann and Patricia West. *They Went to College*, New York: Harcourt, Brace and Co., 1952.

9) Winifred B. Johnson and Lewis M. Terman. "Some Highlights in the

Literature of Psychological Sex Differences Published Since 1920." *J. Psychol.*, 1940, 9, 327–336.

10) A. D. Kinsey, W. B. Pomeroy, C. E. Martin, and P. H. Gebhard. *Sexual behavior in the human female*. Philadelphia: W. B. Saunders. 1953.

11) Frederick H. Lund. "Adolescent Motivation: Sex Differences." *J. Genet. Psychol.*, 1944, 64, 99–103.

12) Jane Ellen McCormick. *Individual factors associated with academic success*. Unpublished master's thesis, Cornell Univ., 1954.

13) John H. McNeeley. *College Student Mortality*. Bull. 11, U. S. Dept. Interior, Office of Educ., U. S. Govt. Printing Office, Washington, 1938.

14) V. R. Mankad. "Infectious Mononeucleosis: An Analysis of 399 Hospital Cases," *Student Med.*, In press.

15) Dorothy Schaffter and Thomas Woody. "Education of Women," in *Enclyclopedia of educational research*, Walter S. Monroe (Ed.), New York: The MacMillan Co., 1950.

16) Amram Scheinfeld. *Women and Men*. New York: Harcourt, Brace and Co., 1944.

17) Social Science Research Center, Cornell Univ. *Report on the Cornell student body*, Mimeographed, 1951.

18) John Summerskill. "Health Services in American Colleges and Universities: Some Questions Answered." *Student Med.*, In press.

19) Martha Margaret Thompson. *Admission information as predictors for graduation*. Unpublished master's thesis, Cornell Univ., 1953.

20) Ruth G. Weintraub and Ruth E. Sally. "Graduation Prospects of an Entering Freshman." *J. ed. Res.*, 1945, 39, 116–126.

THE SOCIAL STRUCTURE OF THE CLASSROOM GROUP: AN OBSERVATIONAL FRAMEWORK

GALE E. JENSEN

Institute of Education, Grinnell College

For roughly forty years we have been studying the instructional designs used by or "on" classroom groups to determine their effectiveness as measured by individual achievement and group productivity. Some of these studies have recognized indirectly that the social structure of the classroom in some way influences group performance and individual learning.[1] Some, through the use of sociometric techniques, have studied certain aspects of the social structure directly and have pointed out its influence upon classroom performance and learning.[2] However, little has been done to provide a framework of concepts that could be utilized to formulate and analyze systematically the classroom problems that are in some way related to the social structure of the classroom group.[3] The purpose of this paper is to outline such a conceptual framework.

This framework, it is hoped, will represent a step toward recognizing that the systematic and conscious utilization of conceptual schemes in educational practice is more powerful for dealing with educational problems than are present "common sense" approaches. The framework is not to be taken as "true" or as the only one inferable from existing research data, but rather accepted as something which, while tentative, is necessary for the orderly examination and description of problems of social structure in the classroom. This, of course, means that the framework is subject to reconstruction whenever more adequate concepts can be formulated.

---

[1] Cf. W. S. Monroe, "Methods of Teaching," *Encyclopedia of Educational Research*. New York: The Macmillan Co., 1950, pp. 745–752.

[2] H. H. Jennings, *Sociometry in Group Relations*. Washington, D. C.; American Council on Education, 1948.

[3] Cf. Ronald Lippitt, "Applying New Knowledge about Human Relations in Educational Practice," *Baltimore Bulletin of Education*, March–June, 1952, pp. 23–30.

THE CONCEPTUAL FRAMEWORK

The form which a description of the social structure of the classroom group takes should be such that it eventually can be integrated with a descriptive framework for examining the dynamics of the classroom. An account, therefore, of the structural aspects of the classroom group should be stated in terms of functional dimensions[4] along which members relate to one another in their attempts (1) to maintain an optimum balance between possible gratifications and deprivations of individual personality needs, and (2) to participate as members concerned with problems both external to the group. Members keep changing their relationships to one another, but always, it is proposed, the changes are along one or more of the dimensions that will be identified. In brief, the dimensions identified indicate all the possible types of functional relationships that individuals can establish between one another *as group members*.[5]

As revealed by the research references cited throughout this article, the dimensions of the social structure of the classroom group represent manifestations of *needs* class members have (a) as individual personalities and (b) as group members. To participate effectively in the tasks or work of the group and to meet their individual needs, members are required to relate to other members in ways that make these accomplishments possible. Stated otherwise, each group member has *need* to establish between himself and the other group members, the kind of relationships that are required for members (1) to work effectively as a group and (2) to maintain a satisfactory balance between the possible gratifications and deprivations of personal needs. These are the basic conditions for a collection of individuals to become and remain a classroom group.

The dimensions along which members seek to relate to one another in a classroom group are seven in number. These are (1) the problem-solving dimension, (2) the authority-leadership dimension, (3) the power dimension, (4) the friendship dimension, (5) the personal prestige dimension, (6) the sex dimension, and (7) the privilege dimension.

---

[4] For a more complete statement of the "functional dimensions approach" to social psychology, cf. Jay M. Jackson, "Analysis of Interpersonal Relations in a Formal Organization," unpublished doctoral dissertation, University of Michigan, 1953, pp. 13–17. To be published as a Research Center for Group Dynamics monograph. The theoretical development which follows has been influenced by this work.

[5] Cf. Dorwin Cartwright and Alvin Zander. *Group Dynamics: Theory and Research*. Evanston, Illinois: Row, Peterson and Company, 1953, p. 417.

The Problem-Solving Dimension.—There are two kinds of problems to which the classroom must continuously devote its energies.[6] One kind pertains to the efforts which the group will expend to organize relations with its external environment. This constitutes the public (or publicly recognized) task of the group. The other kind of problem refers to the efforts of the group to create the kind of intra-group member relationships (its internal socio-psychological environment) that will satisfy members to such a degree that the group will have sufficient energies for working on problems arising from relationships between itself and its external environment. These two problems, common to all classroom groups, can best be understood through descriptions of the environmental settings to which they are indigenous.

The *external environment* of the classroom group can be depicted as a field of forces that impinge upon the group and require it to expend some of its time and energies coping with the "outside." These forces at any given time may be psychological, economic, political, psychical, or sociological in character.[7] They are the result of conditions existing between two distinct parts of the external environment of the classroom group. These two parts are identifiable as follows. First, the "school" as a social system encompasses or incorporates the classroom group.[8] As a social system, the "school" represents one part of the external environment of the classroom group. Second, the classroom group is basically a training group for enabling people to become competent for participating in the affairs of their physical, biological, and social worlds. Therefore, the other part of the external environment to which the classroom group must relate is that which lies outside the school and which we generally identify as the "community." The tasks, projects, and problems upon which the group works are to a large extent determined by and selected from this part of the external environment. The content of the school curriculum is composed of our systematized knowledges and perceptions of the phenomena and problems that exist in the "community." These systematized knowledges and perceptions (cultural artifacts) about community phenomena and problems are referred to as school-subjects or disciplines.[9] They are utilized by the school as instructional resources or means for achieving the public goals of the school, *viz.*, to teach school participants (students) to live intelligently and satisfyingly in the community.

---

[6] Herbert Thelen, "Educational Dynamics: Theory and Research," *Journal of Social Issues*, Volume 6, No. 2, 1950, pp. 31–32.

[7] Cf. Clyde B. Moore and William E. Cole. *Sociology in Educational Practice*. New York: Houghton Mifflin, 1952, pp. 33–54.

[8] Gale E. Jensen, "The School as a Social System," *Educational Research Bulletin*, Ohio State University, February, 1954, pp. 38–46.

The *internal environment* of the classroom group can be presented as the inter-member relations operating within the classroom. The group must expend some of its energies in examining the relationships to be organized and rebuilt from time to time between group members. Because these inter-member relations affect the needs systems of individual group members, the problems they create within the classroom group will usually take priority over any others.[10] Little work on problems induced by forces of the external environment will be accomplished when internal difficulties are numerous and strong.

If the classroom group is to contribute effectively to the realization of the public goals of the school, it must deal skillfully with the two ever-present or persisting problems that arise in its internal and external environments. Members of the classroom group must develop perceptions and expectations about what is required in one another's behavior for deriving adequate solutions to the problems. They must make certain that the functions attached to "student" and "teacher" rôles[11] for purposes of dealing effectively with group achievement and group maintenance are activated at the right time. In brief, individuals have, as members of the classroom group, a *need* to relate to one another along this dimension or the class will be unable to establish the group stability and unity required for productive class work and individual development. Failure of class members to relate to one another along this dimension frustrate the strong personal needs of some members to learn more about the area of study for which the class is responsible. Under such conditions, these members are likely to withdraw or engage in non-constructive behavior.

---

[9] Clyde B. Moore and William E. Cole, *Op. Cit.*, pp. 330–331.

[10] N. T. Fouriezos, M. L. Hutt, and H. Guetzkow, "Self-Oriented Needs in Discussion Groups." In D. Cartwright and A. Zander, *Op. Cit.*, pp. 354–360. Max Corey, "Interpersonal Relations and the Work of the School," *Baltimore Bulletin of Education*, March–June, 1952, pp. 16–22.

[11] Cf. Kenneth Benne and Paul Sheats, "Functional Roles of Group Members," *Journal of Social Issues*, Spring, 1948, pp. 42–47.

In summary, the ways in which classroom group members relate to one another for working on problems induced by the forces of the external and internal environments of the group, represents the problem-solving dimension of the group. The question which reveals the status of this dimension at any one time is: Who engages in or carries on what problem-solving functions with whom to accomplish the public task of the classroom group?

The Authority-Leadership Dimension.—As a classroom group works on its tasks, it is confronted periodically with having to make decisions that affect its progress. The above discussion referring to the problem-solving dimension of the classroom group indicates that the effectiveness with which these decisions are likely to be made is dependent upon having the problem-solving functions activated at the "right" time. To make certain these problem-solving functions are performed at the right time, the group must appoint and/or recognize some members as having the responsibility for and the power to facilitate the problem-solving conditions requisite to making periodic decisions affecting group progress.[12] The status of this authority-leadership dimension in a class for a given time is revealed by the following question. Who is recognized by whom as having the responsibility and power for facilitating the problem-solving conditions necessary for effective decision-making by the class?

By way of example, if the task on which a class is working should change radically, the class will need to consider changes in on-going and contemplated action. In most instances, one finds that some member of the class is recognized as the responsible person for bringing the class together and calling its attention to certain difficulties with which it may be confronted. This member may carry the designation of "teacher", "chairman", or "leader". He is recognized by class members as the "group convener". He has the power to reconvene the class at the proper time and in a place known to all members.

There are numerous other authority-leadership functions that must be performed to insure the conditions necessary for and conducive to effective decision-making by a group. These functions must be identified by the class and then assigned to members who will assume the responsibility for facilitating their performance at the crucial points of class operation. Along with identifying and assigning the responsibilities for performing authority-leadership functions, the class must identify and assign, either explicitly or implicitly, the prerogatives to be used in carrying out these functions.[13]

---

[12] Herbert Thelen, "The Experimental Method of Classroom Leadership," *The Elementary School Journal*, October, 1952, pp. 76–85.

To summarize, the members of the classroom group have a *need* to recognize certain members as having the authority and responsibility for facilitating the problem-solving conditions required for effective decision-making at critical points of class work. In terms of the social structure of the classroom group, this means that class members must establish between themselves a network of authority-leader relationships commensurate with the character of the goals and problems with which the group is confronted.

The Power Dimension.—During the operations of the classroom group it becomes apparent that some members have greater ability (1) to influence the direction the class may take and (2) to induce other members to change their thinking and attitudes. The development of these power relationships in the classroom group is inevitable because members can meet their respective needs only through participation in social situations in which varying needs and resources are matched in ways satisfactory to the participating parties.[14] The character of the relationships along this dimension at any given point in class operation can be discovered by asking: Who possesses what means for gratifying or depriving whose needs?

At times this greater ability to influence others may rest with those who are recognized as having certain authorities. At other times the greater degree of influence is held by those who held no official authority positions. But whether or not this ability to influence rests with those who hold official authority positions is of little importance for the discussion of this dimension. The significant thing is that class members with greater influence are perceived at any given time to have greater resources for gratifying the needs of the group and of individual members.[15] It is this ability to gratify or deprive the needs of others that represents the power an individual member possesses in the class. It is this power as manifested in his ability to gratify or deprive needs that enables a member at any given time to influence the direction the group will take or to induce another member to change his perceptions and attitudes.[16]

---

[13] Herbert Thelen, "Engineering Research in Curriculum Building," *Journal of Educational Research*, April, 1948, p. 582.

[14] Herbert Thelen, "Basic Concepts in Human Dynamics," *Journal of the National Association of Deans of Women*, March, 1952, p. 104.

This power to gratify and deprive needs may be derived from the crucial position a member holds within the group, or it may be derived from his superior ability as an individual member to perform certain acts, control resources, or solve problems. A particular position in the social structure of the class, particular skills, and particular resources, however, are secondary for determining the power a class member holds. Basically, it is the character of the relationships between his social position, resources and skills, and the needs of other members of the class that determines the power a member may have.

To summarize briefly, the way in which class members align themselves with one another for fulfilling individual and group needs through complementary usage of one another's abilities, social positions, and resources represents the power dimension of the classroom group.

The Friendship Dimension.—In the various situations in which class members find themselves, each develops perceptions and feelings about his ability to insure a satisfactory gratification-deprivation balance. Periodic assessments of group situations are made by individual members to determine the possible gratifications and/or deprivations of needs. These assessments are made particularly when the class is trying to project action and/or come to a decision.[17]

---

[15] Ronald Lippitt, Norman Polansky, Fritz Redl, and Sidney Rosen, "The Dynamics of Power." In D. Cartwright and A. Zander, *Op. Cit.*, pp. 480–481.

[16] J. I. Hurwitz, A. Zander, and Barnad Hymovitch, "Some Effects of Power on the Relations among Group Members." In D. Cartwright and A. Zander, *Op. Cit.*, pp. 491–492.

[17] Herbert Thelen, "Educational Dynamics," *Op. Cit.*, pp. 19–21, and H. H. Jennings, *Leadership and Isolation*. New York: Longman's, Green & Co., 1950, chapter 1.

Whenever a situation is perceived or felt to be one which will likely produce an unfavorable gratification-deprivation balance, a group member experiences anxiety and becomes reluctant to express publicly many of his perceptions and feelings. These anxiety feelings may be either consciously or unconsciously experienced. But whether the behavior portrayed is conscious or unconscious in character, many of the perceptions and feelings developed in classroom group situations remain as privately-held experiences by individual members.[18]

At times the class as a whole may be willing to entertain and attempt to deal with individual members' feelings and anxieties; but even when this is true, the group members concerned very often feel that their privately-held perceptions and feelings should not be made public because they will be too threatening to themselves, other members, or the class as a whole.

As a means of allaying anxieties and testing perceptions and feelings being privately held, class members seek to share these perceptions and feelings with other members in whom they feel they can confide. Stated otherwise, classroom group members have a *need* to express and test their privately-held perceptions and feelings with members who will respond actively and supportingly to them. These members are most often those designated as "friends", the intimate associates of a clique or coterie to which they belong.[19]

These relationships which class members seek to develop with one another in order to share and test private perceptions and feelings resulting from participation in classroom situations can be identified as the "friendship" dimension. The question which reveals the status of this dimension at any given time is: Who in the class shares private feelings and ideas with whom?

The Personal Prestige Dimension.—In the process of a classroom group's setting goals for carrying out the learning tasks set by the school curriculum, individual members make various kinds of contributions to the group's progress and welfare. In this way, the group discovers the talents and special worth of its different members.[20]

---

[18] Cf. Raymond L. Gorden, "Interaction between Attitude and the Definition of the Situation in the Expression of Opinion." In D. Cartwright and A. Zander, *Op. Cit.*

[19] Provisions for meeting these needs are sometimes institutionalized in a school system in the form of a counselling service.

Each member comes to be valued to some degree by the group. Each is accorded personal worth commensurate with his talents, personal characteristics, and contributions to the projects of the group. This acquisition of personal worth represents the group's approval of a member's contributions and the acceptance of him as a person with membership rights, privileges, and responsibilities. For a given period of class operation the status of these relationships can be established by determining who is valued by whom by virtue of contributions made to the welfare of the group and the accomplishment of its goals.

The general approval of him by the group along with the special prestige symbols with which he may be rewarded constitutes the source of one of the most important kinds of psychological rewards a class member receives. "To be wanted", "to belong", "to be approved", or "to win recognition and be honored" is one kind of gratification for which it is common for all people to strive. Without a minimum of this kind of gratification or reward, an individual class member's behavior becomes subject to disorganization, unpredictability, and ineffective problem orientation. Without it a member finds it difficult to continue to take part in the life of the class. In brief, individuals as members of a class have a *need* for being valued by the group.

Briefly stated, the way in which members are related to one another in terms of their being valued by the other group members represents the personal prestige dimension of the classroom group.

The Sex Dimension.—There are two sets of forces that operate so as to make the members of a classroom group relate to one another sexually. First, the society in which members of a classroom group are born and live ascribes or assigns to them different rôles that are based in part upon their sex.[21] People formulate different expectations of one another based on differences in sex and in the classroom these expectations will influence the way in which class members will relate to one another. In almost every kind of situation, formal and informal, in which males and females participate, there are socially defined, common expectations as to (1) what activities are appropriate for the different sexes, (2) how females should behave in the presence of males, (3) how females should behave toward other females, (4) how males should behave when with females, (5) how males should perceive and react to other males. These expectations are brought to the classroom and utilized by members, mostly unconsciously, to relate to one another.

---

[20] Cf. C. Heinicke, and Robert Bales, "Developmental Trends in the Structure of Small Groups," *Sociometry*, February, 1953, pp. 7–38.

[21] Kingsley Davis. *Human Society*. New York: The Macmillan Company, 1950, pp. 98–103.

Second, members of a classroom group relate to one another in respect to sex because of certain physiological and psychological forces that affect them as individual organisms.[22] That is, each member has physiological and psychological sex needs which move him or her to seek out other class members who have complementary sex needs and/or who might gratify these needs. In everyday life this relationship is often identified as "love".

The relationships class members endeavor to develop between one another (1) to develop group rôles that take account of culturally defined differences in sex and (2) to satisfy psychological and physiological sex drives represent the sex dimension of the classroom group. The status of the relationships along this dimension at any time is determined by establishing (1) which class members are attempting to define participation in the group according to differences in sex, and (2) which members are relating to one another for purposes of gratifying sex needs?

The Privilege Dimension.—From the first of its history, the classroom group includes at least one member, who because of the rôle he or she holds, receives gratifications (or rewards) in the form of deferences, benefits, advantages, and gifts. The classroom group incorporates this one rôle from the very first of its life because the rôle is assigned to it by the larger social organization of which it is a part, *viz.*, the school. This rôle is that of "teacher".

While the person who is competent to fill the position of "teacher" doubtless makes valuable contributions to the work of the group and achieves personal prestige because of these contributions, he also receives some gratifications by virtue of merely filling the rôle. In brief, the teacher from the first of the group's existence is eligible to receive deferences, benefits, and gifts from members.

---

[22] P. Symonds. *The Dynamics of Human Adjustment.* New York: D. Appleton Century Co., 1946, pp. 31–35.

By virtue of holding the position of "teacher", he has certain privileges that no other member has.

As a class has common experiences, builds a culture for itself, and develops a work structure for effective problem-solving, provisions may be made for rewarding members on a privilege basis in addition to those based on their contributions to the achievement of class goals. The group creates and designates rôles that make members holding them eligible for gratifications in the form of advantages, deferences, and gifts. It seems as though individuals, as group members, have a *need* to insure themselves of a constant receipt of gratifications, especially during those times when they are unable or do not have the opportunity to acquire rewards through making contributions to the work of the class. This attempt to control the "flow" of gratifications seems to represent a sort of "hedging" operation by which members attempt to build into the social structure of the class a sort of "guaranteed set of rewards". The more a group builds a "surplus" of needs-gratifying conditions, the more likely class members will seek to work out privilege relationships between one another. The status of the privilege relationships in a classroom group is determined by discovering who has the right to what gratifications by virtue of the rôle they hold rather than by virtue of the contributions they make to the class goals.

Privilege gratifications may be attached to certain rôles in order to insure that they will be filled by persons who are especially competent in making contributions to the work of the group. They may be assigned to a rôle in recognition of the contributions some group member who presently holds it has made in the past to the welfare and needs of the group. Or these gratifications may be assigned to the rôle as a way of "pointing" to the authority figures of the group. Whatever the reasons may be for attaching these gratifications to certain rôles of the group, the point to be recognized and emphasized is that whatever member fills them becomes eligible for receiving the privilege rewards attached to them.

APPLICATION OF THE FRAMEWORK TO CLASSROOM GROUP OPERATION

In the above description of the problem-solving dimension, it was noted that there are two general problems with which classroom groups are continuously confronted. One has to do with building a kind of class organization that will insure high productivity in the sense of (a) effectively accomplishing the tasks set for the class by the school curriculum, and (b) maximizing the growth or achievement of individual members. The other has to do with maintaining the kind of intra-group conditions that enable members to expend their energies upon class work rather than upon the resolution of inter-member or inter-personal difficulties. These two problems of classroom group operation can be identified as the problems of *group productivity* and *group cohesiveness.*

Class productivity and individual achievement decrease whenever either of these two persisting problems remain unresolved for any length of time. The problem which constantly confronts the teacher is that of diagnosing or analyzing the class to determine (1) which of the two general problems besets the class at any given time and (2) what the character of the relationships are that produce the problem.

Without some kind of conceptual framework to indicate "what kind of things to look for" and "how to look at them", it is not likely that the teacher will be able to bring about a resolution of problems induced by the social structure of the class. When this is the case, class productivity and individual achievement falls and the group more or less "bashes" about with little direction given to its efforts. When confronted with these conditions, a teacher tends to feel extremely threatened and is likely to resort to dictatorial or seductive methods in order to gain more security as the main classroom authority. Under these conditions, class members, including the teacher, seldom understand what is happening to them and it is not surprising that they become highly anxious, reject class work, and some become hostile while others take flight or establish extreme dependency relationships with the teacher.

The systematic use of a framework, such as has been outlined, makes it possible for the educational practitioner to avoid being victimized by the situation just described. For example, the conceptual framework of this article identifies for the educational practitioner the dimensions along which members of a classroom group have a *need* to relate to one another. These dimensions of member relationships constitute the factors relevant to problems induced by the social structure of the classroom group. These different kinds of relationships are "what practitioners must look at" to discover what the conditions are that create at a given time either one or both of the two general problems of classroom group operation. The conceptual framework also indicates to the practitioner that the productivity of a class, the achievement of its members as individuals, and the member-satisfaction with life in the class is a function of the relationships within and between the seven different dimensions. This is "how" a classroom group must be conceived in order to analyze or diagnose effectively the conditions that upset class productivity and class cohesiveness.

SUMMARY

This article outlines a seven-dimensional framework that can be used to analyze the two general problems of classroom operation with which teachers are constantly confronted. These problems were identified as class productivity and class cohesiveness.

The seven dimensions contained in the framework represent the different kinds of relationships that members of a classroom group have *need* to establish between themselves. They also identify, for the educational practitioner, the factors relevant to operational problems induced by the social structure of the classroom group. In addition, the framework indicates to the practitioner that the productivity of the class, the growth of its members, and general member-satisfaction with life in the class must be conceived as functions of the member relationships within and between the seven identified dimensions.

The conceptual framework enables the practitioner to recognize that whenever the member relationships along and between these dimensions become (1) inadequate for accomplishing the work of the class as set by the school curriculum and (2) unsatisfactory for meeting member's needs, the class will probably experience a drop in productivity and individual achievement. The most effective group arrangement for restoring a high level of productivity is to systematically utilize a conceptual framework (1) to locate the dimensions containing the inadequate or unsatisfactory relationships and (2) for determining how these relationships are affecting class productivity and class cohesiveness.

MENTAL MATURITY VERSUS PERCEPTION ABILITIES IN PRIMARY READING

SISTER MARY JAMES HARRINGTON, S.C.L.

Saint Mary College, Xavier, Kansas

and

DONALD D. DURRELL

Boston University, Boston, Massachusetts

There appears to be great confidence in mental age as a basis for beginning reading. It is often assumed that a mental age of six is the minimum for success in reading, and that later progress is closely allied to mental age.

On the other hand, there is growing evidence that success in reading rests more on specific backgrounds of perceptual abilities, and that these are relatively independent of mental age. Among the background abilities that have been shown to be related to success in reading are the following: visual and auditory perception of word elements (*1*, *2*, *3*, *6*), and ability in phonics (*5*, *6*, *7*, *8*)

It seemed desirable to study the relative influence of mental age and the various perceptual factors on reading, using a single population, and evaluating each factor free from the effects of the others. Such a study should provide light on the following questions: How much of reading success depends upon:

1) Visual discrimination of word elements, required for "look-and-say" reading?

2) Auditory discrimination of word elements, which provides the basis for phonetic reading?

3) Phonics—the attachment of sounds to letters and blends, now coming again into favor.

4) Mental age, the general ability commonly used in educational prognosis?

The second grade at the end of the year seems the most appropriate place to evaluate these factors. Reliable measures were available for the various abilities, the intelligence test used required no reading, and the reading vocabulary was small enough for fairly precise measurement. At this stage, too, school instruction plays a relatively large part in reading growth as compared to independent reading.

Reading achievement was measured by two tests: an oral reading test and a silent reading word classification test. Both were measures of reading vocabulary rather than comprehension, since the factors studied are presumed to be related to growth in sight vocabulary. The oral reading test was a story which contained 136 unrepeated words from the basal text used by the pupils, *The Cathedral Basic Readers*. The reliability of this measure, obtained by the split-half method and corrected by the Spearman-Brown Formula, was .98. The word classification test was built upon words not necessarily in the basal readers, but in the Dolch list of "First Thousand Words of Children's Reading." It required the child to read the word silently and put it into the proper classification of meaning. This test correlated .80 with the oral reading test.

Visual discrimination was evaluated by showing the children a word on a flash card and having them identify it from memory by drawing a circle around it in a multiple-choice situation. The reliability of this test, computed by the split half-method and corrected by the Spearman-Brown formula was .80. The auditory discrimination test measured the child's ability to notice initial consonant sounds, rhyming at the ends of words, final consonants, and a combination of initial and final consonants in words spoken by the examiner. This test, designed and evaluated by Nason (4), required the child to circle words to indicate his knowledge of the stimulus sounds. The split-half, corrected reliability of this test was .90. Ability in phonics was measured by an individual test which asked the child to give sounds of letters, consonant blends, and to show a knowledge of the influence of the final silent "e" on word pronunciation. This test had a reliability of .98, obtained by the split-half method and corrected in the usual manner.

Mental age was measured by the *Otis Quick Scoring Mental Ability Test, Alpha*, Form A. This is largely a measure of ability to understand oral language and to follow directions given by the examiner.

All tests were administered by the first author, and were given to five hundred children in the parochial schools in the Archdiocese of Boston. The children represented a wide range of economic status. All of the schools gave training in auditory and visual discrimina-

TABLE I—RESULTS OF MATCHING 161 PAIRS OF CHILDREN TO DETERMINE EFFECT OF PHONICS ABILITY ON READING ACHIEVEMENT

Variable	Mean of High Group	Mean of Low Group	Difference of Means	S.E.$_{diff}$	Critical Ratio
Phonics	45.30	17.96	27.34	1.03	26.54
Mental age	97.93	97.46	.47	1.38	.34
Auditory	27.12	26.79	.33	.73	.45
Visual	22.38	21.89	.49	.39	1.24

tion of word elements, and all had a formal or semi-formal program in phonics.

The primary technique for analysis of the influence of each of the factors independent of the others was that of pairing children who were nearly equal on each of three variables, but markedly different in the variable being studied. When the influence of phonics on reading was being studied, for example, 161 pairs of children were found who were equal in mental age, in visual discrimination and auditory discrimination, but one child in each pair was superior to the other in phonics. The technique of pairing always required the children to be within one-half standard deviation on the controlled factors, and more than one standard deviation apart on the independent factor.

Table I illustrates the results of pairing children to determine the effect of phonic ability on reading achievement. This table is typical of the pairings for the various factors. The means of the controlled variables seldom showed a difference of one point of raw score. The difference of the means of the variable being studied was always large, usually more than one and one-half standard deviation.

If the factor being studied has a high influence on reading, the higher children in the pairing should have a higher mean score in reading than the low group. If, however, the factor is of little importance in reading, the two groups should show approximately equal scores in reading.

Table II shows the results of the pairings and the influence of each factor on reading achievement. The factor being studied is indicated by the pairing data being placed in italics.

All three perceptual factors were found to be significantly related

TABLE II—RELATION OF VARIOUS FACTORS TO READING ACHIEVEMENT

Variable	No. of Pairs	Auditory		Visual		Phonics		Mental Age		Reading			
										Scores		Diff.	C.R.
		High	Low	High	Low	High	Low	High	Low	High	Low		
Auditory	106	*34*	*24*	20	20	34	33	93	93	192	174	18	2.7
Visual	134	31	31	*34*	*17*	34	34	95	95	205	173	32	5.8
Phonics	79	28	28	19	19	*38*	*11*	94	94	180	147	33	4.1
Mental age	107	31	31	21	21	31	31	*105*	*87*	192	189	3	.5

Figures in italics indicate the factor being compared to reading.

to reading achievement. Auditory perception made a difference of 18 words in mean reading score; visual perception, a difference of 32 words; phonics, a difference of 33 words. Apparently each has a significant place in the word analysis program, and each makes a distinct and different contribution to progress in developing a sight vocabulary.

Mental age, however, had little influence on success in reading achievement. The high group had a mean mental age of eight years and nine months as compared to seven years three months for the low group; the two groups were equal in the other perceptual abilities; the difference of a year and a half in mental age produced only a difference of three words in reading vocabulary.

A possible explanation for this lack of influence of mental age may be found in the fact that the mental test used is primarily a measure of oral language comprehension. This is the major element measured in most primary grade intelligence tests. It is likely that even the dullest children who come to first grade have a speaking and listening vocabulary adequate for success in first grade reading, and that additional language fluency does not greatly enhance the rate of acquiring a reading vocabulary in primary grades. The minimum estimates of speaking and listening vocabularies of beginning first graders are 2,500 words or above. Since only 500 words are taught in the first grade, and since the more difficult basal readers rarely present more than 1,800 words in the first three grades, it should not be surprising that the relationship between most current primary grade mental tests and reading achievement is low.

A partial correlation study using the entire population of five hundred children confirmed the findings of the paired comparison study, with the minor exception that visual discrimination showed a higher correlation with reading than did phonics. The zero-order correlations with readings for the four factors were as follows: visual discrimination, .64; phonics, .56; auditory discrimination, .54; mental age, .23. Partial correlations were as follows: visual discrimination, .45; phonics, .32; auditory discrimination, .26; mental age, .000.

Since the findings in regard to mental age and phonics are of particular interest, it seemed desirable to repeat the study on a larger population with a wider geographic distribution. Through the coöperation of Sister Mary Brendon, RSM, of Arkansas, Sister Mary Hortense, RSM, of Oklahoma, Sister Catherine Marie, SCL, and Sister Ann Lorraine, SCL, both of Kansas, the study was repeated in the several states.

The same tests and statistical procedures were used with a population of 1,000 children at the end of the second grade. The larger number permitted pairings more than twice the size of the initial Boston study. The findings of this larger study are presented in Table III.

In the data from the midwestern states, the factors relate to reading in the same order as in the Boston study. The one marked difference is in the influence of visual perception, where the difference between the high and low group was only six points, resulting in a somewhat lower influence of the factor in the reading score. The midwestern mental ages ran somewhat higher, permitting a greater difference between the high and low pairs in matching.

TABLE III—RELATION OF VARIOUS FACTORS TO READING ACHIEVEMENT (MIDWESTERN STUDY)

Variable	No. of Pairs	Auditory		Visual		Phonics		Mental Age		Reading			
		High	Low	High	Low	High	Low	High	Low	High	Low	Diff.	C.R.
Auditory	265	*36*	*24*	24	23	40	39	102	101	206	185	21	6.4
Visual	340	32	31	*27*	*21*	41	39	104	104	207	184	23	7.4
Phonics	161	27	27	22	22	*45*	*18*	98	97	189	160	29	6.1
Mental age	283	32	31	24	24	40	40	*122*	*91*	204	198	6	1.6

Figures in italics indicate the factor being compared to reading.

Here a difference of thirty-one months of mental age, or two years and seven months, made a difference of only six words in the mean reading scores.

Referring to the questions stated at the beginning of this article, the findings of these studies are as follows:

1) Auditory and visual discrimination of word elements have high importance in success in acquiring a primary grade reading vocabulary.

2) Phonics instruction is clearly important, having a higher relation to reading achievement than any of the factors studied.

3) Mental age, as measured by the test used, has little influence on success in learning to read.

It is clear from the nature of the pairing that many children who are high in mental ability score no higher than those of low mental ability in the background factors that make for success in reading. This study indicates strongly that specific instruction in phonics and in visual and auditory perception of word elements is essential to success in building reading vocabulary in primary grades.

REFERENCES

1) B. Alice Crossley. *An Evaluation of the Effect of Lantern Slides on Auditory and Visual Discrimination of Word Elements*. Doctor's Thesis, Boston University School of Education, Boston, 1948. (Available on microfilm.)

2) Donald D. Durrell and Helen A. Murphy. "The Auditory Discrimination Factor in Reading Readiness and Reading Disability." *Education* 73: 556–560; May 1953.

3) Helen A. Murphy. *An Evaluation of the Effect of Specific Training in Auditory and Visual Discrimination on Beginning Reading*. Doctor's Thesis, Boston University School of Education, Boston 1943. (Available on microfilm.)

4) Doris E. Nason. *The Influence of Vocabulary Common to Test and Textbook on Primary Reading Test Scores*. Doctor's Thesis, Boston University School of Education, Boston, 1951. Unpublished.

5) Sister M. Dorothy Browne, O. P., *Phonics as a Basis of Improvement in Reading*. Washington, D. C.: The Catholic University of America, 1939.

6) Sister Mary Nila, O.S.F. "Foundations of a Successful Reading Program." *Education* 73: 543–555; May 1953.

7) Joseph Tiffin and Mary McKinnis. "Phonic Ability: Its Measurement and Relation to Reading Ability." *School and Society* 51: 190–192; February 1940.

8) Frank Wilson and others. "Reading Progress in Primary Grades." *Elementary School Journal* 38: 442–449; February 1938.

BOOK REVIEWS

MARGARET NAUMBERG. *Psychoneurotic Art: Its Function in Psychotherapy.* New York: Grune & Stratton, Inc., 1953, pp. 148 $6.75

At the core of Miss Naumberg's latest volume on the use of spontaneous art production in psychotherapy is a three-year study and record of a psychoneurotic patient. A consideration of the psychotherapeutic significance of the spontaneous art productions of this patient takes up by far the largest portion of this volume. Also considered at length are the conclusions drawn from this study. The volume also contains two shorter sections of related material: (1) "The Correlation of the Patient's Rorschach and Other Tests with the Patient's Art Productions" by Adolf G. Woltmann; and (2) "Comments on the Rorschach Test and the Psychotherapeutic Use of Graphic Art" by Zygmunt A. Pietrowski. The book is introduced very favorably with a short preface by Dr. Kenneth E. Appel. The book is an expansion of Miss Naumberg's studies in this field. Of these, the volume on *Schizophrenic Art: Its Meaning in Psychotherapy* has been previously reviewed in this JOURNAL.

The volume includes forty-six illustrations, eight of which are color plates, which show various phases of development as interpreted by the use of these spontaneous drawings. Illustrative of the phases considered are, escape into a pregenital world of fantasy, the family constellation, patterns concerning fear, the growth and development of the patient's ego, the threat of seduction by the father, exploring the meaning of sexuality, marriage and its adjustments, and the patient resolves her fantasy life.

There is also an appendix in two parts: in the first is reviewed the history and development of art in therapy; and in the second is considered the recent literature on clinical and theoretical aspects of spontaneous art in therapy and diagnosis. Some of this material has appeared in previous publications by the author. The material presented, more so than her previous publications, represents a rough measure of a method for evaluating the progress in therapy—ways of detecting, revealing and measuring changes in the patient's personality. Appel in his preface praises Naumberg's therapy as successfully demonstrating the effectiveness of therapy in the area he considers to be refractive; namely, obsessive compulsive conditions. In this particular patient's case, these are compulsive eating, masturbation and marked phantasy life.

As in previous publications, the material here will be useful to therapists using projective techniques in diagnosis as well as therapy. As various publications of Naumberg have appeared, it seems that enough has been written to warrant her writing one more compact volume, tying together the meanings as well as the description of the process of therapy by the use of art productions which is her contribution.

H. MELTZER

Psychological Service Center, St. Louis, Mo.

HENRY CLAY LINDGREN. *Mental Health in Education.* New York: Henry Holt and Co., 1954, pp. 561.

The preface outlines the contents of the book, indicating the plan of organization, and pointing out some things which the author wishes to emphasize. "The chief concern in this book is with the normal child, the normal child who has a few problems (and who has not?), but particularly the normal child who has a healthy need to progress and to develop in ways that are socially and emotionally mature." (p. iv).

The subject matter as shown in the table of contents includes the following in the order given: the mental hygiene attitude in education; development of the child as an individual; motivation; emotional maturity; behavior problems; socioeconomic factors underlying children's behavior; communication; the child and the group; forces operating in children's groups; patterns of acceptance and rejection; three approaches to teaching; the teacher as a person of power and authority; relationship of the teacher to the classroom group; relations with parents; disintegrative influences in education; integrative influences in education; the school as a therapeutic environment; evaluation and diagnosis; the teacher; problems of adjustment. There are five appendixes and an index.

Although 'Mental Health' is stressed in the title, the book is primarily a discussion of problems of education, especially those of teachers. It is essentially a pedagogical text dealing with 'techniques' of teachers, and the 'rôles' of the teachers. Some philosophy of education is brought in to help in the discussions of the various attitudes that teachers may take. These discussions are primarily from the point of view of mental hygiene or mental health. There are many references at the ends of chapters. They seem to be primarily to pedagogical literature; secondarily to mental hygiene publications.

There are many valuable discussions, especially for the beginning teacher. Many of them seem superficial and too brief, but the wide scope of the book may have forced some of this. The emphasis is on the therapeutic, with the usual lack of recognition upon the very important problem of prevention. Do we not look to the schools to build up the attitudes, habits, interests, etc., that will be the greatest bulwark against the development of abnormal and undesirable character and personality? 'Prevention' does not appear in the index, and does not seem to be essential to the thinking in the text. Perhaps the nearest to it is found in the chapter on Integrative Influences in Education, and in relation to Guidance.

The difficulties of teachers are brought out, but there is much of encouragement for those who are looking forward to teaching as a profession. The status of teachers is better than formerly and indeed quite high in the scale. The salaries are much better. Social recognition has greatly improved. After noting current advantages of teaching, the author says: "Probably all of these factors operate to make teaching a more enjoyable, rewarding, and generally worth-while profession than it was a decade ago." (p. 532).

The index is generally well made (pp. 550–561). The volume follows the consensus of large numbers of educators at the present time. It will probably find a very considerable use as a text. There are a sufficient number of broad generalizations which may be questioned, and thus offer opportunity for extended discussions.

A. S. Edwards

The University of Georgia

Lee J. Cronbach. *Educational Psychology*. New York: Harcourt, Brace and Co., 1954, pp. 628.

During the period in which this book was in preparation, the general expectancy among instructors in the relevant courses was that Cronbach would provide us with a notable text. He has. The author's catholic competency in both psychology and education, his willingness to eschew the widely used formula for writing educational psychology textbooks, and the very evident responsibility with which the book was planned and executed have combined to produce one of the best of the approximately twenty texts which presently petition the attention of instructors.

The focal worth of this publication can best be suggested by reference to the author's Preface. Here, in a most provocative statement, Cronbach asserts 'four qualities' that an educational psychology must possess "in order to fill the needs of the present day." These are the objectives for the present book and are announced as: "(1) An educational psychology must be relevant to school problems, (2) An educational psychology must be soundly based on research, (3) An educational psychology must be complex enough to do justice to human behavior, (4) An educational psychology must be clear." The self-imposed criteria presented in prefatory pages too often have only a coincidental relationship to the final product; only the contentious will quarrel with the observation that, in this case, the objectives constitute an adequate summation of the text.

The nineteen chapters are organized under five Parts: "Psychology and school problems," "Readiness and its development," "Acquiring ideas, attitudes and skills," "Planning, motivation, and evaluation," and "Emotional learning." Each of these major sections ends in an extensive summary. The end-chapter 'reading lists' are annotated and inviting. Liberal and meaningful use is made of line drawings, figures, and tables. The author's selection of photographs is to be commended; most of the few included appear to extend textual denotations, clarifications, and suggestions.

The technique of including frequent intra-chapter questions, which Cronbach utilized in his *Essentials of Psychological Testing*, is continued here. Inasmuch as one of the reviewers of the earlier text was skeptical of the value of this device, it may be appropriate to note that students evaluating the text at the reviewer's institution are firmly positive on this point. ARTHUR P. COLADARCI

Stanford University

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume 46 November, 1955 Number 7

LEADERSHIP BEHAVIOR ASSOCIATED WITH THE ADMINISTRATIVE REPUTATION OF COLLEGE DEPARTMENTS

JOHN K. HEMPHILL

The Ohio State University

I. INTRODUCTION

The effectiveness of administration within educational institutions is rarely assessed in a systematic or formal manner. Consequently, in research on problems of educational administration, the staff of the Ohio State Leadership Studies[1] has looked for methods for making reasonable estimates of the administrative effectiveness. This paper is concerned with characteristics of the "reputation for being well administered" as a criterion of the quality of administration in college departments.

Opinions of his peers, co-workers or "buddies" about the quality of an individual's performance have been demonstrated to provide useful criterion data (*11*). The question of whether analogous opinions can be found that will reflect the quality of educational administration is suggested immediately. Are departments of a college characterized by a reputation for their administrative excellence? If such reputations exist do they reflect valid differences in administrative practices? One part of a more comprehensive study of the departments within a single Liberal Arts college provides some evidence to support the proposition that the "reputation for being well administered" is useful criterion information.

---

[1] The Ohio State Leadership Studies, initiated in 1946 by Carroll L. Shartle, have been planned as a ten-year program of research on leadership and administration problems in military, business, industrial, educational, and civilian governmental organizations. The research reported here was supported by a grant received from the Rockefeller Foundation for the study of administration in education. Melvin Seeman and Andrew W. Halpin, as members of the staff of the Ohio State Leadership Studies, have conducted related studies of public school administration.

II. PROCEDURE

The Study: This report concerns only a portion of a larger study of the leadership and administration of 22 college departments in a moderately large mid-western university.[2] The larger study had as its general purpose, the testing of the applicability of research methods being developed by the Ohio State Leadership Studies to the university situation. These methods consisted of series of questionnaires and rating forms which had been used previously in military (*3*), business (*1*) and public school systems (*8*).

Coöperation necessary for the study was first secured from the dean of the Liberal Arts college and the 22 department heads. Later all faculty members were informed of the nature and objectives of the study by means of (a) a letter of introduction signed by the dean, (b) personal meetings with each department's faculty and (c) written instructions for completing a packet of questionnaires to be submitted anonymously. Although participation was voluntary, 234 of the 322 faculty members returned their questionnaires after two follow-up letters.

The questionnaires were distributed in four types of packets made up as follows:

Type A: (These questionnaires were distributed to the 22 department heads)

1) Background Information Questionnaire
2) Leader Behavior Description Questionnaire (self description directions)
3) Responsibility, Authority and Delegation scales (*9*)
4) Reputational Ranking Form

Type B: (These questionnaires were distributed to a randomly selected third of the members of each department's faculty)

1) Background Information Questionnaire
2) Leader Behavior Description Questionnaire (*General* form with directions to describe the actual behavior of the faculty member's department chairman)

---

[2] Vernon Jon Bentz, Psychologist, Sears, Roebuck and Co. assisted in planning this study, and collected the data.

3) Group Dimensions Description Questionnaire
4) Reputational Ranking Form

Type C: (These questionnaires were distributed to a second randomly selected third of the members of each department)

1) Background Information Questionnaire
2) Leader Behavior Description Questionnaire (*General* form with directions to describe the behavior of an "ideal" leader of the faculty member's department)
3) Leader Behavior Description Questionnaire (*Specific* form, with directions to describe the actual behavior of the faculty member's department chairman)
4) Reputational Ranking Form

Type D: (These questionnaires were distributed to a third randomly selected third of the members of each department's faculty)

1) Background Information Questionnaire
2) Status Position Scales[3]
3) Status Perception Scales[3]
4) Group Dimensions Questionnaire
5) Reputational Ranking Form

The four-way division of the work of completing the large battery of forms and questionnaries was necessary in order to reduce the formidability of the requests made of each faculty member.

Only the Background Information Questionnaire, the Group Dimensions, the Leader Behavior Description Questionnaire, (General Form) and the Reputational Ranking Form are of immediate interest in this paper. Each of these four questionnaires is described briefly in the following paragraphs.

Background Information Questionnaire: This form secured data concerning the respondent's department affiliation, age, sex, academic rank, degree held, length of time in his department, length of association as a member of an academic faculty, publications, affiliation with other professional organizations and general job satisfaction.

The Leader Behavior Description Questionnaire: This questionnaire is the product of the joint efforts of the staff of the Ohio State Leadership Studies. Its development is described by Halpin (*3*), and Fleishman (*2*). The form used in this study consisted of

---

[3] Seeman (*7*) (*8*) for a description of the use of these scales.

150 items; however, only 30 of this number were actually scored. This decision was based upon an extensive factor analysis of all the items. Two scores are obtained: (a) Consideration and (b) Initiating Structure.

Consideration refers to leader behavior that is characterized by warm friendly relations with group members and concern with group member welfare. The following items are typical of this dimension:

He looks out for the personal welfare of individual members.
He treats all members as his equal.
He puts suggestions by the group into operation.
He gets group approval on important matters before going ahead.

Initiating Structure refers to leader activities that introduce organization, new ways of doing things, and new procedures for solving group problems, etc. The following are typical Initiating Structure items:

He lets members know what is expected of them.
He encourages the use of uniform procedures.
He asks that members follow organizational lines.
He tries out his new ideas in the group.

The respondent indicates the frequency with which his leader engages in each of these behaviors by selecting one of five adverbs: always, often, occasionally, seldom or never.

The Group Dimension Questionnaire: The development of this questionnaire is described in detail by Hemphill and Westie (*5*). It yields scores on 13 dimensions or characteristics of groups:

1) *Autonomy* is the degree to which a group functions independently of other groups and occupies an independent position in society (13 items).

2) *Control* is the degree to which a group regulates the behavior of individuals while they are functioning as group members (12 items).

3) *Flexibility* is the degree to which a group's activities are marked by informal procedures rather than by adherence to established procedures (13 items).

4) *Hedonic Tone* is the degree to which group membership is accompanied by a general feeling of pleasantness or agreeableness (5 items).

5) *Homogeneity* is the degree to which members of a group are

similar with respect to socially relevant characteristics (15 items).

6) *Intimacy* is the degree to which members of a group are mutually acquainted with one another and are familiar with the more personal details of one another's lives (13 items).

7) *Participation* is the degree to which members of a group apply equal effort to group activities (10 items).

8) *Permeability* is the degree to which a group permits ready access to membership (13 items).

9) *Polarization* is the degree to which a group is oriented and works toward a single goal which is clear and specific to all members (12 items).

10) *Potency* is the degree to which a group has primary significance for its members (15 items).

11) *Stability* is the degree to which the group resists changes in its size and in turnover of its members (5 items).

12) *Stratification* is the degree to which a group orders its members into status hierarchies (12 items).

13. *Viscidity* is the degree to which members of the group function as a unit (15 items).

The respondents (faculty members) describe their departments by indicating how well the items assigned to each dimension characterize their respective departments as groups.

The Reputational Ranking Form: This form is based on the methodology of the "nominating technique" (*6*), (*10*). The form was prefaced with these directions:

"It is desirable to ascertain an estimate of departmental 'reputation'. *Not including your own department*, rank the five departments in the Liberal Arts College that have general reputations on the campus for being 'Best Led' or 'Best Administered' departments.

"You are not asked to give you own personal evaluation, but asked to give the general reputation of the department, using as your point of reference what 'most people in the College of Liberal Arts would agree' that a given department has as a given reputation. For example, you may personally think a certain department is poorly led but know that the reputation of that department on the campus is generally higher than your own personal opinion would rate it. In making the ranking use this general campus reputation."

Similar instructions also were given with a request for the five departments that were reputed to be "Least Well Led" or "Least

Well Administered." With completion of the form the respondent had indicated his knowledge of the campus reputations of ten departments and by implication had placed the 12 remaining departments in a category not distinguished by reputations of being either well administered or poorly administered.

Analysis of Data

The analysis of data was planned in a manner to provide tentative answers to many questions that are raised when reputational information is suggested as a criterion of quality of administration. Among these questions are:

1) Is there any degree of consistency or agreement among the individuals who report knowledge of reputations of the different departments?

2) Are these reputational rankings made by individuals qualified to know existing campus opinions?

3) What characteristics of a department or of the behavior of its chairman are reflected in the department's campus reputation with respect to quality of administration?

Scoring of the Reputational Rankings

A relatively arbitrary scoring procedure was adopted for the purpose of summarizing the data secured by the reputational ranking forms. Each nomination received by a department was assigned the weight according to the following scheme:

Nomination	Weight
First Best Administered	5
Second Best Administered	4
Third Best Administered	3
Fourth Best Administered	2
Fifth Best Administered	1
First Less Well Administered	−5
Second Less Well Administered	−4
Third Less Well Administered	−3
Fourth Less Well Administered	−2
Fifth Less Well Administered	−1
Not nominated for either category	0

The score derived for each department was the sum of the weights assigned to the nominations that were received. Scores

TABLE I—DEPARTMENTAL ADMINISTRATION "REPUTATION SCORES" DERIVED FROM TWO MATCHED SAMPLES OF RESPONDENTS AND FROM THE TOTAL NUMBER OF RESPONDENTS

Sample A (N = 33)		Sample B (N = 33)		Total Sample (N = 133)	
Dept.	Scores	Dept.	Scores	Dept.	Scores
A	9	A	2	A	16
B	−29	B	−30	B	−83
C	47	C	80	C	203
D	−46	D	−40	D	−112
E	49	E	73	E	223
F	−9	F	−18	F	−14
G	−53	G	−37	G	−138
H	16	H	9	H	31
I	45	I	19	I	100
J	−41	J	−54	J	−122
K	89	K	72	K	252
L	−8	L	−5	L	−24
M	0	M	0	M	4
N	101	N	68	N	305
O	−35	O	−26	O	−87
P	−74	P	−70	P	−271
Q	−1	Q	9	Q	13
R	−29	R	−12	R	−53
S	−27	S	−31	S	−80
T	19	T	−2	T	25
U	−17	U	−12	U	−56
V	3	V	1	V	3

ranged from −271 to 305. The distribution of the scores for the 22 departments is given in the third column of Table 1.

III. RESULTS

The presentation of the results of the analysis of the data will be organized in respect to the three major questions that were asked about the characteristics of "reputation for being well administered" as a criterion of administrative effectiveness.

What is the degree of consistency or agreement among the reputations reported for the various departments? This question was answered by establishing two samples of respondents and determining the agreement between the two sets of reputation scores derived from the two samples. The two samples of respondents included only those who had made relatively complete nominations (8, 9 or 10, of the five best administered and the five less well administered) and was constituted so as to have an approximately equal number of respondents from the various departments in each sample. Table 1 shows the "reputation scores" of the department as derived from the two samples. The correlation between the two sets (Sample A vs. Sample B) of "reputations" scores is .94. If the Spearman Brown prophecy formula is applied to this coefficient, we may estimate the reliability of these reputation scores as .97.

We, therefore, can have considerable confidence in the existence of a campus reputation with respect to the quality of departmental administration. Evidently, this reputation is known and can be reported with a high degree of agreement among members of the college faculty.

What are the characteristics of those faculty members who completed the Reputational Ranking Form? Not all of the faculty members who coöperated in the study by returning questionnaires completed the Reputational Ranking Form. Many faculty members indicated on the face of the form that they were unaware of campus opinion about the administrative reputation of departments. Others made only a portion of the ten nominations requested. It was possible to establish two groups of respondents on the basis

TABLE 2—AGE OF RESPONDENTS COMPOSING (*a*) A GROUP OF SEVENTY-THREE WHO RETURNED THE REPUTATIONAL RANKING FORM COMPLETED AND (*b*) A GROUP OF ONE HUNDRED AND ONE WHO RETURNED THE FORM BLANK

Age	Returned Completed Forms (N = 73)	Returned Blank Forms (N = 101)
Over 70	2	2
65–70	5	2
60–64	6	4
55–59	8	8
50–54	16	8
45–49	14	6
40–44	10	26
35–39	7	18
30–34	5	26
25–29	0	0
Under 25	0	1

TABLE 3—LENGTH OF TIME A MEMBER OF THE COLLEGE FACULTY OF RESPONDENTS COMPOSING (*a*) A GROUP OF 73 WHO RETURNED COMPLETED FORMS AND (*b*) A GROUP OF 101 WHO RETURNED BLANK FORMS

Length of Faculty Membership	Returned Completed Forms (N = 73)	Returned Blank Forms (N = 101)
Over 6 years	44	25
Four to 6 years	5	5
Two to 4 years	15	31
One to 2 years	7	21
Six months to 1 year	2	15
Less than 6 months	0	4

of their responses to the Reputational Ranking Form. Group I consisted of 73 faculty members who provided at least eight of the ten nominations requested. Group II consisted of 101 faculty members who completed the other questionnaires in their packets but failed to make a single nomination on the Reputational Ranking Form. A comparison of data available on the characteristics of these two groups of respondents indicated that those faculty members who provided nominations differed from those who did not in that they were considerably older and, what is perhaps more important, had been members of the faculty for a longer time. (Tables 2 and 3).

From Table 2 we note that 51% of the respondents who returned completed forms were over 50 years of age. This compares with 24% of the respondents who returned blank forms.[4]

Referring to Table 3, we see that 60% of the respondents who returned completed forms had been members of the faculty 6 years or more. This compares with 25% of the respondents who returned blank questionnaires.[5]

It is to be expected that campus reputations of college departments are known through long associations of faculty members. Individuals who are relatively "new" on the campus are likely to be unaware of events that were significant in establishing a department's reputation with respect to its administrative excellence.

---

[4] The difference between the means of the two groups, (47.8 and 40.5) is significant at the .01 level (t = 4.8).

[5] The difference between the means of the two groups (4.30 and 2.72) is significant at the .01 level (t = 4.7). The first three categories of Table 3 were combined to make this test.

We interpret the tendency of more mature (older) and better qualified (a member of the faculty for a longer length of time) faculty members to complete this form to indicate a face validity of the information we obtained. We also must accept the fact that judging campus reputation of departments may be a meaningless and impossible task for a large portion of the younger faculty.

Is the reputation of departments for being well administered related to the style of the department chairman's leadership? The packets of questionnaires that were distributed to the members of each department included the Leader Behavior Description Questionnaire, General Form. One third of the members of each department completed this questionnaire to provide their descriptions of the leadership behavior of the chairman of their respective departments. A second third of the members of each department completed this same questionnaire but with instructions to indicate what an "ideal" chairman for their department should do. For 18 of the 22 departments descriptions were available of the "actual" and the "ideal" behavior of the chairmen from at least 2 and as many as 17 department members. For each of these 18 departments, an average Consideration score and an average Initiating Structure score were calculated. Table 4 presents the average scores for both "actual" and "ideal" leadership behavior and the "discrepancy" between these sets of scores.

Pearson product-moment correlation coefficients were computed to express the relationship between each of the two dimensions of actual leadership behavior and the "reputation" score of the departments. Correlations were also computed to express the relationship between the absolute discrepancy (either positive or negative) between "actual" and "ideal" leadership behavior on the two dimensions and the "reputation" score. These data are presented in Table 5.

By reference to Table 5 it can be seen that there is some relationship between the style of leadership of the department chairman as this is viewed by department members and the department reputation on the campus for being well or poorly administered. This is especially true if we view the departure of the actual behavior of the department from what is held by members of his department to be ideal behavior. Departments that achieve a reputation for good administration are those departments led by chairmen who attend to both of the facets of leadership measured

TABLE 4—AVERAGE SCORES ON CONSIDERATION AND INITIATING STRUCTURE OF DEPARTMENT CHAIRMEN AS DESCRIBED BY THE DEPARTMENT MEMBERS

Department	Consideration			Initiating Structure		
	Ideal	Actual	Dis-crepancy	Ideal	Actual	Dis-crepancy
A*	44.0	45.3	1.3	44.7	42.7	2.0
B	51.0	42.0	9.0	46.5	30.7	15.8
C*	49.3	48.1	1.2	45.5	43.5	2.0
D	43.0	24.3	18.7	34.3	33.0	1.3
E*	46.4	42.1	4.3	36.9	36.7	.2
F	45.5	38.3	7.2	44.8	44.5	.3
G*	53.0	35.0	18.0	50.0	26.5	13.5
I*	47.8	49.6	1.8	41.8	49.2	7.4
J	45.5	49.5	4.0	42.0	25.5	16.5
K*	47.3	41.1	6.2	38.4	38.9	.5
L	47.3	40.0	7.3	42.7	32.5	10.2
N*	49.0	44.8	4.2	39.6	37.8	1.8
O	49.3	43.7	5.6	47.0	44.3	2.7
Q*	48.0	47.0	1.0	36.3	39.0	2.7
R	50.3	38.3	12.0	40.8	34.0	6.8
S	50.0	34.0	16.0	40.5	25.0	15.5
T*	46.6	49.2	2.6	39.7	40.2	.5
U	51.0	43.2	7.8	38.8	35.8	3.0

* Above the median on "reputation."

in this study, i.e., they concern themselves with (1) organizing departmental activities and initiating new ways of solving departmental problems, and at the same time with (2) developing warm considerate relationship with members of the department.[6]

If we examine the leadership patterns of department chairmen closely we note a conjunctive relationship between Consideration and Initiating Structure suggesting that a minimal amount of both types of behavior are required for the acquisition of a good reputation and that an excess of one type of behavior does not compensate for the lack of the other. An excess of Consideration behavior cannot compensate for a deficiency in Initiating Structure and vice versa. If a cutting score of 41 on Consideration and 36 on Initiating Structure are applied simultaneously, the reputation

---

[6] This finding is congruent with the results of a similar study conducted within the Air Force and reported by Halpin (*3*).

TABLE 5—CORRELATION COEFFICIENTS EXPRESSING THE RELATIONSHIP BETWEEN THE ADMINISTRATIVE "REPUTATION" SCORES OF 18 COLLEGE DEPARTMENTS AND FOUR LEADERSHIP SCORES OF THE DEPARTMENT CHAIRMEN AS DESCRIBED BY DEPARTMENT MEMBERS

Score	Mean	S.D.	Correlation with Reputation Score
Reputation	21.7	133.8	—
"Actual" Consideration	42.2	6.4	.36
Discrepancy between "Actual" and "Ideal" Consideration	6.9	5.6	−.52*
"Actual" Initiating Structure	36.8	6.7	.48*
Discrepancy between "Actual" and "Ideal" Initiating Structure	6.3	6.9	−.55*

* A correlation coefficient of .47 is significant at the .05 level.

of the department in terms of being above or below the median are successfully identified in 16 out of 18 cases. (These cutting scores were selected by inspection of the data and may have no specific significance.) The data for this comparison are given in Table 6.

The Chi squared value for the data of Table 6 is 8.00, significant beyond the .01 level of probability.

A further means of examining the conjunctive relation between these two facets of leadership behavior is available for the discrepancy score. The total of the two discrepancy scores of department chairmen correlates −.63 (.01 level of probability) with the

TABLE 6—THE RELATIONSHIP BETWEEN THE REPUTATION ACHIEVED BY COLLEGE DEPARTMENTS AND THE CONSIDERATION AND INITIATING STRUCTURE SCORES OF DEPARTMENT CHAIRMEN TAKEN CONJUNCTIVELY (N = 18)

Chairman's Leadership	Above Median Reputation	Below Median Reputation
	Departments	
Score of 41 or larger on Consideration *and* a score of 36 or more on Initiating Structure	A, C, E, I, K, N, Q, T	O
Score of less than 41 on Consideration *or* a score of less than 36 on Initiating Structure	F	B, D, G, J, L, R, S, U

reputation scores. These last two reported results suggest strongly that if department chairmen meet the expectations of their faculties with respect to *both* Consideration and Initiating Structure the department is more likely to achieve a favorable reputation on the campus.

In the remainder of the presentation of results we shall examine characteristics of the departments for relationship with their administrative reputation.

Are the characteristics of their department as seen by department members related to the reputation of their department? The Group Dimensions Questionnaire was distributed to two thirds of each department's members. The questionnaire provides a score on each of 13 group characteristics (see page 388). This score was calculated as the average of the descriptions supplied by those department members who completed this questionnaire. Average scores based on a minimum of two member's descriptions were available for 18 of the 22 departments. Pearson correlation coefficients were computed for each of the 13 dimensions and the reputation score. The correlations are presented in Table 7.

None of the correlations reported in Table 7 is statistically significant. There appears to be no relationship between the manner in which department members characterized their respective de-

Table 7—Correlation between Each of Thirteen Group Dimensions (Average Scores) and Departments "Reputation" for Being Well Administered (N = 18)

Correlation: Reputation vs.	Mean	S.D.	r*
Autonomy	34.4	4.0	.09
Control	28.7	3.4	−.07
Flexibility	37.6	5.8	.02
Hedonic Tone	18.8	2.1	.16
Homogeneity	33.1	3.7	−.05
Intimacy	48.4	4.4	−.02
Participation	46.4	4.5	.29
Permeability	22.5	4.3	−.04
Polarization	38.4	5.6	.05
Potency	51.1	4.6	.05
Stability	15.1	2.8	−.06
Stratification	40.7	5.1	.05
Viscidity	35.5	8.1	.04

* An r of .47 is required for significance at the .05 level.

partments by use of the Group Description Questionnaire and the "reputation" of the departments for being well administered. This lack of relationship highlights the significance of the earlier reported relationship between the behavior of the department chairman and administrative reputation. It points to chairman behavior specifically, rather than general impressions about the department as the more likely determiner of reputation.

Are there relationships between "demographic" characteristics of departments and their reputation? Examination was made of each of the following "demographic" characteristics of the 22 departments and their reputation scores:

1. Age of faculty members (Average)
2. Size of the department (Number of members)
3. Length of service to the department of it's faculty members (Average number of years)
4. Length of academic experience of each department faculty (Average number of years)
5. Academic rank of members (Per cent Associate Professors or Professors)
6. Educational attainment of faculty of the departments (Per cent with Ph.D.)
7. Publications record of department members (Index based on number and "quality" of publications)

Only one of these seven "demographic" variables showed a relationship with reputation. It was found that large departments tended to receive higher reputation scores (r between number of members and reputation = .58, significant at the .05 level). Two hypotheses were advanced in an effort to account for this relation. *First*, it was suggested that larger departments might require more Initiating Structure behavior on the part of their chairmen and since Initiating Structure had been shown to be related to "reputation" the relationship with department size could be a function of the same underlying variables.[7] The *second* explanation involved the suggestion that the significance of administration in larger as compared with smaller departments is well recognized by those responsible for the selection of department heads and that in the case of larger departments more care had been exercised. In other words, in very small departments, chairmen might be appointed on the basis of seniority without much regard for other qualifications, but in large departments demonstration of ability to assume administrative responsibility would be involved in a much larger measure in the selection of a chairman.

---

[7] In previous research a relation between size of group and leader centered direction had been observed (*4*).

The first of the two hypotheses could be checked in part by determining the relation between the size of the department and the Initiating Structure behavior of their chairman. This correlation was computed and found to be .11, which is not significant. The available evidence rejects the first interpretation.

IV. SUMMARY AND DISCUSSION

The campus reputations of 22 departments for being well administrated were secured from members of the faculty of a Liberal Arts college by use of a nominating form. As parts of the same research study data were also available concerning (a) the leadership behavior of department chairmen as viewed and as wanted by department members, (b) group characteristics of departments as viewed by department members, and (c) demographic characteristics of the departments. An examination was made of the relationship among these data in an effort to assess the usefulness of campus opinion as a criterion of quality of administration. The following conclusions resulted from the analysis:

1) Administrative reputation of the college department was reliably reported by faculty members. Agreement between two independent samples of respondents was very high ($r = .94$).

2) Older and more mature faculty members provided a larger proportion of the reputation information than "younger" or "new" members of the faculty. In fact, it appears many new members of the college faculty were not aware of the departments' reputations.

3) Reputation for being well administered is related to the leadership behavior of department chairmen as this behavior is described by department members. Those departments with best reputations for good administration have chairmen who are described as above the average on *both* Consideration and Initiating Structure and who more nearly meet the behavior expected of an ideal chairman. It appears that optimal amounts of both of these types of behavior are required in order that a department earn a reputation for good administration.

4) With the exception of size, all group characteristics of the departments, both demographic characteristics and those de-

scribed by means of the Group Description Questionnaire are unrelated to reputation for good administration.

5) Larger departments tend to have better administrative reputations than smaller departments. This fact is independent of the Initiating Structure activity of the chairman and may indicate only that more care is exercised in selecting chairmen of large departments.

In conclusion, it appears that reputation may provide a criterion of excellence of administration. Many questions remain to be investigated concerning the limits within which such reputation may be useful. Certainly, qualified respondents must be found who have had opportunity to become acquainted with events and opinions bearing on quality of administration. However, in many organizations a pool of qualified respondents may be available in the senior members of the groups. Our study involved only one college faculty. We know nothing about how differences between organizations may affect the availability and quality of opinion regarding administrative reputation. However, in view of the dearth of techniques available for assessing quality of administrative performance and importance of the problem, the suggestion of the usefulness of the reputation criterion reported here should be followed up.

REFERENCES

1) C. G. Browne. A study of executive leadership in business: I. The R. A. and D. Scales. *J. appl. Psychol.*, 1949, *33*, 521–526.

2) E. A. Fleishman. Leadership climate, human relations training and supervisory behavior. *Personnel Psychol.*, 1953, *6*, 205–222.

3) A. W. Halpin. The leadership behavior and combat performance of airplane commanders. *J. abnorm. soc. Psychol.*, 1954, *49*, 19–22.

4) J. K. Hemphill. Group factors in leadership: I. Relations between the size of the group and the behavior of "superior leaders." *J. soc. Psychol.*, 1950, *32*, 11–22.

5) J. K. Hemphill & C. M. Westie. The measurement of group dimensions. *J. Psychol.*, 1950, *29*, 325–342.

6) J. G. Jenkins. The nominating technique its uses and limitations. *Amer. J. Psychol.*, 1947, *2*, 433 (Abstract).

7) M. Seeman. Some status correlates of leadership. In A. G. Grace (Ed.) *Leadership in American Education*. Chicago: University Chicago Press, 1950.

8) M. Seeman. Role conflict and ambivalence in leadership. *Amer. Soc. Rev.*, 1953, *18*, 373–380.

9) R. M. Stogdill & C. L. Shartle. Methods of determining patterns of leadership behavior in relation to organization structure and objectives. *J. appl. Psychol.*, 1948, *32*, 286–291.

10) C. L. Vaughn. The nominating technique. In G. A. Kelly (Ed.) *New Methods in Applied Psychology*. College Park, Maryland: University of Maryland, 1947.

11) R. J. Wherry & D. Fryer. Buddy ratings: Popularity contest or leadership criteria? *Sociometry*, 1949, *12*, 179–190.

THE DEVELOPMENT OF SHORTER AND MORE USEFUL SELECTION TESTS

A. PEMBERTON JOHNSON

Testing and Guidance Division, Newark College of Engineering, Newark, N. J., formerly of Educational Testing Service, Princeton, N. J.

This article describes what has been done within the last few years to shorten two selection tests of proved validity. Any one might have followed a very similar procedure for shortening either of these tests. The technique followed with the second test is, I believe, somewhat new. The two tests are: (1) the Pre-Engineering Inventory shortened to become the Pre-Engineering Ability Test and (2) the 6-hour Law School Admission Test shortened to its present 3½ hour version.

SHORTENING THE PRE-ENGINEERING INVENTORY

The Pre-Engineering Inventory (PEI) was first used in 1943, by the Measurement and Guidance Project in Engineering Education, then conducted by the Graduate Record Office of the Carnegie Foundation for the Advancement of Teaching. Designed primarily as a guidance instrument, which could be used for selection, the PEI provided seven sub-test scores and a composite of three of those sub-test scores. The composite was intended to provide a single index most predictive of 1st term engineering college averages and was quite effective for that purpose. By July 1949 when I became Project Director of the Measurement and Guidance Project evidence had begun to accumulate that many of the users found this test too expensive and too time-consuming, both in the 6-hour Long form and the 4-hour Short form. During 1949–50 a number of discussions were held with users to define the desirable features of a shortened test. In general the users wanted a test as valid as the Pre-Engineering Inventory or nearly so but requiring no more than 90 minutes testing time, a test which was reliable enough for individual guidance yet inexpensive and easy both to administer and to score. The users wanted, also, a means of converting scores on the new test to scores on the Pre-Engineering Inventory, so that their past experience with the earlier test would not be lost. One user wanted two scores: one on comprehension of scientific reading and one on mathematical ability.

Information available to aid us in revising the test included:

(1) Product-moment validity coefficients of sub-test and of composite scores against first-term grades at 12 engineering colleges

(2) Product-moment intercorrelations of sub-test scores

(3) Sub-test and composite score estimated product-moment reliability coefficients

(4) Item validity coefficients at one large engineering college

(5) Speededness data.

Examination and manipulation of these data showed that raw scores on the two sub-tests, Comprehension of Scientific Materials and General Mathematical Ability, when added, would predict first term engineering college grades almost as well as the Composite score which is derived from these two sub-tests combined with a third one, Technical Verbal Ability. Median validities of these two sub-tests at 12 engineering colleges were: for Comprehension of Scientific Materials, .50; and for General Mathematical Ability, .58. The median intercorrelation was .70. The multiple R for the two in predicting first term grades based on the above data was .59 as compared with .60 for the Composite score based on three parts. Odd-even raw score reliabilities for these sub-tests were: .88 for the 100 Scientific Comprehension items and .89 for the 90 Mathematics items. Time limits were: Scientific Comprehension, 90 minutes, and Mathematics, 70 minutes.

Estimated validity and reliability of single test combining half-length portions of the two sub-tests were: validity .57, reliability .90. These values were believed to justify use of the shortened test for selection and guidance of individual students. Study of the item validity data revealed 10 items in each sub-test with item validities of essentially zero. Study of the speededness data suggested that slightly more time should have been provided for both sub-tests in order to make them essentially power tests.

In view of the above information, it was decided to produce from the 90 better Scientific Comprehension items and the 80 better Mathematics items two essentially half-length experimental forms equivalent as possible, allowing somewhat more time for both item types and retaining each type in a separately timed part. Thus two 80-minute forms comprising 45 Scientific Comprehension items in 45 minutes and 40 Mathematics items in 35 minutes were prepared. They were administered during the September 1950 Orientation week in 1-2 and 2-1 order to random halves of a large entering engineering freshman class, N = 610. The previous year's class had taken the Pre-Engineering Inventory. Both classes took the College Board SAT and Physics Achievement Tests which were used as means of equating scores on the experimental tests to scores on the PEI. One of the experimental forms was published in July 1951 as the Pre-Engineering Ability Test. The manual reports a Kuder-Richardson formula 20 reliability of .90 for 305 engineering freshmen and an estimated average validity of .57 against first term grades. A study at Carnegie Institute of Technology by Dr. Roland C. Moore reports a validity coefficient against first semester weighted averages there of .68 for 260 students to whom the 80-minute Pre-Engineering Ability Test was administered. This figure was no lower than correlations of the PEI against first semester averages for the PEI Composite in the previous years. The findings from use of the test appear to be consistent with the prior theoretical estimates of reliability and validity. Thus it was possible to provide an 80-minute test with a single score virtually as valid as the composite score on the 4- to 6-hour test, together with comparable scores for the earlier and the newer test. Two reliable part scores could not be obtained in 80 minutes with the kind of test material used. I was responsible for the planning and arranging for the execution of this work which was supervised largely by Dr. William H. Angoff of our Statistical Analysis Department and Mr. Bernard S. Cayne of the Test Development Department.

SHORTENING THE LAW SCHOOL ADMISSION TEST

The shortening of the *Law School Admission Test* (LSAT) presented a somewhat different problem. The first two forms of that test comprised 10 to 12 sub-tests, the raw scores of which were summed for direct conversion to the single standard scores reported to users. Test analyses and validity studies of these first two forms showed odd-even coefficients of reliability for these unspeeded aptitude tests of the order of .95–.97, and suggested the possibility of shortening these forms without reducing validity coefficients of reported scores. These validity coefficients against a criterion of first year law grades, ranged upward to .65 with a 12 school mean of .40 for the test as compared with .38 for pre-law averages. (The first year law school grade criterion has a greater judgmental and less mathematical basis than the first year engineering college criterion.)

As a means of providing greater test security, utilizing available items efficiently, and of making greater funds available for the development and tryout of promising new materials, and for other services to users, it was decided to consider how best to develop four 3½-hour tests from the earlier forms and from new pretested materials. Several staff discussions as to the best method of shortening the earlier forms of the LSAT led to plans for an empirical comparison of five selected methods applied to one form. The study was planned and executed under the direction of Dr. William B. Schrader and Dr. Marjorie A. Olsen who reported it at the 1952 meetings of the American Psychological Association. For convenience it dealt with reducing six sub-tests which required 190 minutes testing time to a shortened test of 135 minutes, even though the original and present forms of the LSAT are both longer than these times indicate.

The analysis was based on LSAT total and sub-test scores and on first year law school average grades for two random cross-validation sub-samples of 500 from among 1000 first year students in nine law schools. In order to combine criterion data from all law schools first year grades were converted to a common scale; the regression of grades on LSAT total score for each school was used to determine the appropriate mean and standard deviation for the group on the common criterion scale.

The necessary calculations and item selections were made for each of the five methods with data from each of the two sub-samples so that trial scoring keys could be used for cross-validation on the opposite sub-samples.

Keys based on one sub-sample were then used to rescore the answer sheets for the other sub-sample, and the new trial scores by each method were correlated with the first-year law school grades.

Keys based on three item-analysis methods were tried out:

1) Internal consistency item analysis—items showing the highest biserial correlations with LSAT total score were selected for this trial key.

2) External criterion item analysis—items having the highest biserial correlations with first-year law school grades were included in this key.

3) E/I ratio, substantially that described by Gulliksen in the Proceedings of the 1948 Invitational Conference on Testing Problems—items with the highest ratios of point biserial correlation

with grades (external criterion-E) to point biserial with sub-test score (internal criterion-I).

Keys for the two following methods, each concerned with estimating optimal lengths of time for sub-tests, were also explored:

4) Horst's method for condition III is described in his June 1951 article in *Psychometrika*. This procedure takes into account the validity, reliability and original length of each sub-test and its intercorrelations with other sub-tests. Under the conditions that total time (or number of items) is fixed and the score is total number right, it determines the optimal time for each sub-test so that the validity of the total score is maximized. In this study it was not quite possible for some sub-tests to include the optimal number of items because the number required was greater than the number available.

5) The linear method—In principle, this procedure is similar to Horst's method, but by making certain assumptions, it leads to a markedly simplified solution. It takes into account the validity and original length of each sub-test and its intercorrelations with other sub-tests. It specifies an amount of time for each sub-test in the shortened form such that the raw score regression weights of all of the sub-tests are approximately equal. It assumes that, with moderate changes in the length of any sub-test, standard score regression weights for the sub-test scores change proportionately with length of time for the sub-test. In effect the latter assumption is equivalent to assuming sub-test reliabilities of 1.00.

If these assumptions are accepted, the solution for optimal sub-test times follows readily. For any sub-test, its proportion of the total test time is simply the product of its standard score regression weight and its original time allotment divided by the sum of such products for all sub-tests. Within each sub-test, the items with the highest biserial correlations with the sub-test score are chosen to the extent required to meet the needed sub-test length.

Surprising enough, within each sub-sample, the cross-validity coefficients for the five different methods were virtually the same. In one sample they ranged from .483 to .496, a difference of .013; in the other, from .569 to .576, a difference of .007. The simplest method by far, the linear method, had a slightly higher mean validity coefficient in the two samples than any other method.

The linear method was applied to nine sub-tests in a second form of the LSAT using data for two other random sub-samples of 500 law students each. In the multiple regression equation for these nine sub-tests, four sub-tests showed appreciable beta weights and the remaining five showed negligible weights, when used with the other four. A trial 135-minute key, developed to provide optimal time allotments among these four tests and to use the most internally homogeneous items of each sub-test showed a cross-validity coefficient of .466 as compared with .454 for the 300 minute original. This is a striking example of the effectiveness of the method.

The findings of the linear method applied to the two earlier forms of the LSAT, together with tests and item analysis data from these forms, and a careful inventory of pre-tested and analyzed items of each type made possible the development of a plan of assembly of four alternate 3½-hour forms which maximize the validity of all. The data available strongly suggests that the four new forms will be at least as valid as the earlier long forms. A routine follow-up validity study on the first of the alternate forms is planned, however, for next fall when the first new criterion data should be available. Reliabilities of two of the new forms are .913 and .927. The four new forms include carefully devised sets of items enabling the equating of scores on one pair of the new forms to scores on the first long form and on another pair to those on the second long form. Linking items are provided to link the new forms across pairs.

Planning and execution of the Law School Admission Test work described above has involved a number of the ETS staff, principally, Dr. Ledyard R. Tucker, Drs. Schrader and Olsen of the ETS Statistical Analysis Staff and Dr. Edith Huddleston of the ETS Test Development Department. I have served as Project Director suggesting goals and aiding in planning to achieve the ends mutually agreed upon.

DISCUSSION

You may have noticed that although the details of methodology in shortening the engineering and law tests were different, the basic principles followed were somewhat the same. Initial shortening was by elimination of sub-tests which added negligibly to total test validity. The length of each included sub-test was determined partly or wholly on the basis of multiple regression data. The items required for each sub-test were chosen by some appropriate item selection technique.

A COMPARATIVE STUDY OF VALUES AND CRITICAL THINKING SKILLS OF A GROUP OF EDUCATIONAL WORKERS

JOSEPH C. BLEDSOE

The University of Georgia

INTRODUCTION

Formal school experiences such as college courses frequently have as their objectives the development of certain attitudes, insights, and skills. Such courses sometimes make claims which may not be justified. It is admittedly somewhat difficult to assess some of the more important attitudinal and behavioral changes which may take as a result of such experiences. Among mature students, moreover, changes in attitudes and practices are more difficult to bring about since basic patterns have largely been established. This study reports an attempt to measure, at least partially, certain value-patterns and critical thinking skills of a group of graduate students with particular reference to changes which are revealed following an intensive six-weeks course in research methods. Certain correlates of value scores and critical thinking scores with factors such as age, sex, and area of professional work are also presented.

PURPOSES

Specific purposes of the study, which may be roughly regarded as hypotheses, are outlined as follows:

1) What are the patterns of values and critical thinking skills of this group of graduate students in education? Can a distinctive hierarchy of values for such a group be established?

2) What changes, if any, in dominant life-value are brought about in a period of six weeks among a group of relatively mature educational workers? (The six weeks period is characterized by an intensive course in research methodology.) Does such a course which emphasizes practical or applied research result in a significant increase in interest in theoretical or practical values?

3) What changes, if any, in critical thinking skills are apparent following such a course? Can improvement in these skills be brought about in such a short time?

4) Are there significant sex and age differences in values and in critical thinking skills?

5) Are there significant differences in values and in critical thinking skills among three types of workers, viz., administrators, elementary teachers, secondary teachers?

6) Are there significant correlations between value scores, critical thinking scores, scores on subject-matter tests on research methodology, age and sex?

Brief Description of Group Studied, Course Taught, and Tests Employed. The subject group consisted of 46 graduate students enrolled in a graduate course entitled "Methods of Applied Research in Education." There were 22 men and 24 women; 16 persons were employed during the regular school year as administrators (mostly principals), 10 as elementary school teachers, and 20 as high-school teachers. The group ranged in age from 25 to 51, with a mean age of 36 years and a standard deviation of seven years. In experience, the group ranged from two years to 26 years, with a mean of 11 years. This particular section, a summer-school group, was apparently representative of graduate students in education enrolled at the University of Georgia.

The course in research methods, required of all graduate students majoring in education, has as its broad objectives the development of insights, skills, and competencies needed to become effective producers and consumers of educational research, with special reference to applied research. Among the more detailed objectives are an increased awareness of the need for applying research methods to the solution of problems, a clear conception of applied research as a means of solving problems, some knowledge of the more widely-used methods and tools of research in education, increased competence in interpreting the results of research, and general understanding of what is usually involved in preparing the research report. Among the procedures employed are lecture-discussions, presentation by students of proposed projects, and critiques by instructor and students of studies reported in the literature.

The inventory of values employed was the Allport-Vernon-Lindzey *Study of Values*, Revised (1951) Edition. *The Study of Values* is designed to measure the relative value of six basic interests in personality: the Theoretical, Economic, Aesthetic, Social, Political, and Religious. The classification is based upon Eduard

Spranger's *Types of Men*, a work which assumes that the personalities of men are best known through a study of their values or evaluative attitudes. The scale consists of a number of questions, based upon a variety of familiar situations to which two alternative answers (in Part I) and four alternative answers (in Part II) are provided. In all there are 120 answers, 20 of which refer to each of the six values. Usual time required is approximately 20 minutes.[1]

The instrument employed for determining critical thinking skills was the Watson-Glaser *Critical Thinking Appraisal*. This instrument is designed to provide problems and situations which require the application of some of the important abilities involved in critical thinking. The subtests of the Critical Thinking Appraisal are (1) inference; (2) recognition of assumptions; (3) deduction; (4) interpretation; and (5) evaluation of arguments. There are two forms available, each of which has 99 items; time required is approximately 40 minutes.[2]

In addition to these standardized tests, the author, as instructor of the course, gave a 116-item teacher-made test based on the subject matter of research methodology. This test was of the multiple-choice and alternative-response type, and represented the fifth revision (sixth form) of the test after use with preceding classes.

PROCEDURES

At the first class meeting, Form Am of the Watson-Glaser *Critical Thinking Appraisal* and the Allport-Vernon-Lindzey *Study of Values* were administered in one sitting. Students were later informed, upon their request, in confidential interviews of their individual percentile scores on the Watson-Glaser Test and the order or hierarchy of the individual scores on the Allport-Vernon-Lindzey *Values* inventory. The course proceeded in the usual manner as briefly described in a preceding paragraph. At the end of the six-week period, the subject matter test was given and the two standardized instruments were administered, including the alternate Form Bm of the Watson-Glaser *Critical Thinking Ap-*

---

[1] Gordon W. Allport, Philip E. Vernon, and Gardner Lindzey. *Manual of Directions*, Boston, Houghton Mifflin Company, 1951.

[2] Goodwin Watson and Edward M. Glaser. Manual, Watson-Glaser *Critical Thinking Appraisal*. Yonkers-on-Hudson, New York, World Book Company, 1952.

praisal. Normative data such as sex, age, and area of professional work were recorded. Measures of central tendency, variability, and correlation were then computed on these data and test scores. Tests of significance involving differences in means, standard errors of the differences, and the "t" ratio of significance were employed.

ASSUMPTIONS AND LIMITATIONS

The following assumptions and limitations of the study should be mentioned:

1) It is assumed that the two standardized instruments employed have at least some validity for the purposes of this study. The validity of the findings is limited to a large extent by the validity of the instruments employed.

2) The study was primarily exploratory. No control group was employed to determine what growth or changes may have taken place in activities other than experiences of this particular type; generalization about cause-and-effect relationships is therefore somewhat limited.

3) It is assumed that these value patterns and thinking skills are important outcomes, and that students are sufficiently concerned that they give accurate responses.

4) Most students in the subject group were taking one additional course during this period; this course may have been of the same or of a different type, which may have augmented or detracted from changes produced by the experiences of the research methods course.

SUMMARY OF FINDINGS

Table I presents mean scores for both the beginning and the end-of-course administrations of the two standardized inventories, along with standard deviations and differences in means. Differences in value scores from pre-test to end-test are very small and none of them are significant; largest difference comes in the relatively unstable "social" subscale. The increase in mean score on the *Critical Thinking Appraisal* scale of approximately four points meets the five percent level of confidence test. Since the course had as one of its aims that of an increased awareness of the need for a critical attitude toward problems, it appears reasonable to conclude that students may have sharpened some of these skills.

TABLE I—MEAN SCORES, STANDARD DEVIATIONS, AND DIFFERENCES IN MEANS OF SCORES ON ALLPORT-VERNON-LINDZEY STUDY OF VALUES AND WATSON-GLASER CRITICAL THINKING APPRAISAL

	Mean Scores		Standard Deviations		Differences End-Minus Pre-Test
	Pre-Test	End-Test	Pre-Test	End-Test	
Theoretical	41.20	41.09	6.87	9.43	−.11
Economic	41.09	41.65	7.11	7.39	.56
Aesthetic	35.35	35.39	7.05	7.67	.04
Social	38.11	36.80	5.62	6.30	−1.31
Political	37.70	38.83	5.54	7.40	1.13
Religious	46.41	46.65	8.64	7.28	.24
Critical Thinking Appraisal	63.28	67.32	12.71	10.58	4.04*

* Significant at five per cent level of confidence.

It also appears fairly certain that a short course of this type is not likely to produce a significant change in basic values. It may be, however, that the real impact of such an experience is not immediately felt, and that with increased practice in use of a problem-solving approach, such values may shift more than can immediately be measured. Since no control of outside influences seems feasible, no easy way is suggested to discover whether or not such a conclusion is warranted.

Table II presents differences among the subscales of the Allport-Vernon-Lindzey *Study of Values* for the subject group. This is for the first administration of the test, which, as has been demonstrated above, is not significantly different from the later scores. Inspection of the table reveals a definite hierarchy of values for the group-as-a-whole. Highest value is that of the Religious, the mean score of which is significantly higher than the mean scores of the other five subscales. Second in order of importance and approximately equal are the values Theoretical and Economic. Next in order and approximately equal are the scales Social and Political. Finally and least important to the group-as-a-whole is the Aesthetic value. To determine whether or not this hierarchy of values is representative of teachers-in-general, or is peculiar to this particular group, or perhaps representative of teachers in a Southern state or in the Southern region, is not within the scope of this study. It is interesting to note, however, that, as compared

TABLE II—DIFFERENCES BETWEEN MEAN SCORES AMONG SUBSCALES OF THE ALLPORT-VERNON STUDY OF VALUES

Subscale (N = 46)	Economic	Aesthetic	Social	Political	Religious
Theoretical	.11	5.85**	3.09*	3.50**	−5.21**
Economic		5.74**	2.98*	3.39*	−5.32**
Aesthetic			−2.76*	−2.35	−11.06**
Social				.41	−8.30**
Political					−8.71**

Read the table as follows: The difference in mean scores of the Theoretical scale and the Economic scale is .11. Plus differences are favorable to the subscale listed vertically; minus differences are favorable to the horizontal subscale.

* Significant at 5% level of confidence.

** Significant at 1% level of confidence.

TABLE III—INTERCORRELATIONS BETWEEN SUBSCALE SCORES OF THE ALLPORT-VERNON STUDY OF VALUES

Subscale	Economic	Aesthetic	Social	Political	Religious	Second Administration of Same Subscale (at End of Six Weeks)
Theoretical	−133	−141	−501**	−149	−165	716**
Economic		−325*	−402**	−040	−199	697**
Aesthetic			246	−357*	−392**	782**
Social				−098	−078	435**
Political					−131*	292*
Religious						558**

Read table as follows: Correlation between subscale scores on Theoretical and Economic is −.133, etc. Repeat test reliability after six weeks of Theoretical scale is .716. Decimals have been omitted.

* Significant at 5% level.

** Significant at 1% level.

with national norms, Southerners, particularly Southern women, tend to score high on the Religious subscale.[3]

Table III presents intercorrelations between subscale scores of the Allport-Vernon-Lindzey *Study of Values.* All intercorrelations are negative, with the exception of that between the Aesthetic-Social scales. This correlation of .246 is not, however, significantly different from zero. Highest negative correlation comes in the

---

[3] Susan W. Gray. "A Note on the Values of Southern College Women: White and Negro." *Journal of Social Psychology*, XXXV (1947), 239–241.

TABLE IV—CORRELATIONS BETWEEN SUBSCALE SCORES OF THE ALLPORT-VERNON STUDY OF VALUES, WATSON-GLASER CRITICAL THINKING APPRAISAL SCORES, SUBJECT MATTER TEST SCORES, AND AGE

	Watson-Glaser CTA, Form A	Watson-Glaser CTA, Form B	Research Subject-Matter Test	Age
Theoretical	161	092	048	−138
Economic	−319*	−305*	−256	−007
Aesthetic	−219	−200	023	565**
Social	149	105	368**	−002
Political	242	198	018	−156
Religious	049	117	−059	−194
Watson-Glaser CTA, Form A		604**	636**	−348*
Watson-Glaser CTA, Form B			551**	−366**
Research Subject-Matter Test				−135

* Significant at 5% level.
** Significant at 1% level.
Decimals have been omitted.

relationship between the Theoretical and Social Scales. Correlations significant at the one per cent level include the Theoretical-Social, Aesthetic-Religious, and Economic-Social. Also given in Table III are the correlations between the first and second administrations of the tests which essentially constitute repeat test reliability estimates after a period of six weeks. These reliability coefficients range from a low in the case of the Political Scale of .292 to a high of .782 for the Aesthetic Scale. The mean reliability coefficient using the z transformation was .61. This reliability coefficient is considerably lower than that reported in the Manual of Directions of the Test (.82).

Table IV presents the correlations between the scales of the values inventory, the Watson-Glaser *Critical Thinking Appraisal* scores, scores in the subject matter test on research methods and chronological age. Most of the correlations between the scales of values and the two forms of the Watson-Glaser Tests are insignificant. Only in the instances of the relationship between the Economic scale and the two forms of the *Critical Thinking Appraisal* is the correlation significant. The repeat-test reliability, after six weeks, of the two forms of the Watson-Glaser *Critical Thinking Appraisal*, is .604, which is considerably lower than the coefficients

TABLE V—SEX DIFFERENCES IN MEAN SCORES ON THE SUBSCALES OF THE ALLPORT-VERNON-LINDZEY STUDY OF VALUES AND OTHER VALUES

Variable	Mean Scores		Difference (Male minus Female)
	Male (N = 22)	Female (N = 24)	
Values			
Theoretical	43.13	39.42	3.71
Economic	40.05	42.04	−1.99
Aesthetic	31.68	38.71	−7.03**
Social	37.50	38.67	−1.17
Political	39.41	36.13	3.28
Religious	47.82	45.13	2.69
Watson-Glaser CTA, Form A	66.09	60.71	5.38
Watson-Glaser CTA, Form B	69.36	65.42	3.94
Research Subject-Matter Test	56.36	58.17	−1.81
Age	32.86	39.21	−6.35**

* Significant at 5% level of confidence.
** Significant at 1% level of confidence.

reported in the manual. (These coefficients range from .79 to a .95.) A significant correlation between the scores on the Social subscale and scores on the subject matter test was found, as well as significant correlations between the two forms of the Watson-Glaser *Critical Thinking Appraisal* and the Research Subject Matter Test. A highly significant and somewhat surprising correlation of .565 was found between scores of the Aesthetic scale and age. All other values subscales were negatively and insignificantly correlated with age. Negative correlations of .348 and .366, both significant, were found between the Watson-Glaser *Critical Thinking Appraisal* and age.

Table V presents sex differences in mean scores on the several variables. Both men and women scored highest in the Religious scale and there seemed to be considerable similarity in the hierarchy of values between men and women. The only significant difference was in the instance of the Aesthetic scale in which women scored substantially higher than men. That this group is an atypical one, as compared with the general population, is brought out by the fact that the Religious scale, particularly in the case of men, is much higher than for the general population. In this particular

TABLE VI—MEAN SCORES AND DIFFERENCES IN MEANS ON SUBSCALES OF THE ALLPORT-VERNON STUDY OF VALUES, AND OTHER VARIABLES OF A GROUP OF GRADUATE STUDENTS CLASSIFIED BY AREA OF PROFESSIONAL WORK

Values	Mean Scores			Difference in Means		
	Admin. (N = 10)	Elem. (N = 16)	Secondary (N = 20)	Admin. minus elem.	Admin. minus sec.	Elem. minus sec.
Theoretical	42.4	39.3	41.2	3.08	1.18	−1.90
Economic	37.6	43.1	42.9	−5.54*	−5.34*	.20
Aesthetic	36.5	38.5	32.9	−2.00	3.65	5.65*
Social	39.8	39.0	36.3	.81	3.51	2.70
Political	38.1	36.6	37.9	1.46	.21	−1.25
Religious	45.1	43.8	48.8	1.26	−3.74	−5.00
Watson-Glaser CTA, Form A	62.4	60.9	65.2	1.48	−2.82	−4.30
Watson-Glaser CTA, Form B	67.3	65.8	68.2	1.51	−0.89	−2.40
Research Subject-Matter Test	60.4	56.8	55.1	3.58	5.28	1.70
Age	38.5	39.5	32.7	−1.00	5.85*	6.85**

* Significant at 5% level of confidence.
** Significant at 1% level of confidence.

sample, men scored even higher than women on this scale. For both men and women, Religious, Theoretical and Economic scales were the first three choices, and the Political, Social and Aesthetic were the least preferred. Men scored highest on the *Critical Thinking Appraisal Test* for both forms; however, these differences were not significant. Women scored higher on the subject matter test; again, this difference was not significant. The difference in mean ages between men and women was significant.

Table VI presents mean records and differences in means among the several variables of the group classified by area of professional work. The differences in means among the three groups were insignificant, except for a significant preference for the Economic in the case of the elementary teachers and the secondary teachers, as compared with administrators, and a significant preference for the Aesthetic for the elementary teachers as compared with the secondary teachers. This latter preference may be suggested by the fact that most of the elementary teachers were women and a substantial number of the secondary teachers were men. The differences in age between the administrators and secondary teachers and the elementary teachers and secondary teachers were significant.

SUMMARY

This study has reported mean scores in certain broad categories of values as revealed by responses to the Allport-Vernon-Lindzey *Study of Values* and also mean scores on the Watson-Glaser *Critical Thinking Appraisal* of a group of 46 mature graduate students of the University of Georgia both before and after an intensive course in research methodology.

An analysis of these scores revealed a hierarchy of values for the group with the Religious scale listed as of greatest preference, followed by the Theoretical and Economic (on the same level), the Social and Political (on the same level), and finally the Aesthetic. No significant differences in pre-course and end-course mean scores in values were found. A fairly significant (five per cent level) difference in mean scores on the Critical Thinking Appraisal from pre-course to end-course was obtained. Differences in mean scores for men and women for the values inventory were found to be insignificant except for the Aesthetic scale, in which women made a higher mean score. Other differences (except age) were insignificant.

THE DAVIS-EELLS GAMES AS A MEASURE OF THE INTELLIGENCE OF POOR READERS

JOSEPH JUSTMAN and MIRIAM ARONOW

Bureau of Educational Research, Board of Education of the City of New York

For many years, the New York City schools have used the Pintner Intermediate test as a measure of general intelligence at the sixth grade level. Criticism of the relatively high reading load of this instrument has led the Bureau of Educational Research to undertake a comparative study of several other group measures of intelligence with a view to determining whether a more satisfactory test might be utilized.

One of the tests considered for possible use was the Davis-Eells Games, Elementary. Bureau staff members administered this measure in nine sixth grade classes drawn from six elementary schools. A total of 246 pupils, to whom the Pintner Intermediate Test had previously been administered, were tested. Both tests were administered within the same school month. This paper reports the findings of a comparison between the results obtained on the two instruments.

Character of the Sample.—Because the study was concerned primarily with the appropriateness of the tests for poor readers, the nine classes in which the Davis-Eells Games were administered were selected to constitute a sample somewhat below the average of the sixth grade population at large. The group contained a high proportion of third and fourth grade readers who had difficulty in reading the Pintner Test and whose general capacity, it was felt, was underestimated by the Pintner Test. It was thought that an instrument like the Davis-Eells Games, which required no reading, would yield higher IQ's for the group than the Pintner Test, and that the mean IQ on the Davis-Eells Games would probably lie between 95 and 100.

THE FINDINGS

1) Differences in Mean IQ—As a first step in the analysis of the obtained data, mean IQ's and standard deviations were computed for the two tests. In addition, the product-moment correlation be-

TABLE 1—MEAN IQ'S AND CORRELATION BETWEEN IQ'S ON DAVIS-EELLS GAMES AND PINTNER TEST

Test	N	Mean	SD	r
Davis-Eells	246	89.3	13.3	.50
Pintner	246	88.5	14.4	

tween the IQ's obtained through the administration of the Davis-Eells Games and the Pintner Test was determined. These data are presented in Table 1.[1]

Although the Davis-Eells Games and the Pintner Test show only moderate agreement in the relative ranking of pupils, less than a one-point difference in mean IQ appears between the two tests. This difference is not statistically significant. The mean IQ of below 90 obtained on the Davis-Eells Games does not conform to the expected pattern.

2) Differences in Individual IQ's—The mean scores presented in Table 1 provide a gross comparison. The differences in the IQ's obtained by individual pupils on the Davis-Eells Games and the Pintner Test are summarized in Table 2.

When the two tests are compared, approximately the same proportion of the pupils obtained their higher IQ on the Davis-Eells Games as on the Pintner Test. It should be noted, too, that 63 per cent of the pupils (including 4.1 per cent showing no difference in IQ) obtained IQ's which are within 10 points of each other. On the other hand, 10.6 per cent of the obtained IQ's on the two tests differed by more than 20 points.

3) Differences in IQ as Related to Intelligence Level—A further analysis of the data is summarized in Table 3, which presents the mean difference in IQ between the Davis-Eells Games and the Pintner Test as related to intelligence level on the latter.

A consistent trend is noted in mean IQ differences between the Pintner Test and the Davis-Eells Games as one ascends the Pintner IQ ladder. For pupils obtaining Pintner IQ's between 70 and 79, the mean difference between Pintner IQ and Davis-Eells IQ is 7.8 points, in favor of the latter. In the case of children with Pintner IQ's between 90 and 99, the difference becomes 1.2 points, and is

---

[1] The courtesy of the Division of Test Research and Service of the World Book Company, publishers of the two measures, in providing these data is gratefully acknowledged.

TABLE 2—NUMBER AND PROPORTION OF PUPILS SHOWING DIFFERENCES IN IQ BETWEEN DAVIS-EELLS GAMES AND PINTNER TEST

Difference	Higher on Pintner		Higher on Davis-Eells	
	N	Per cent	N	Per cent
1–10 points	80	32.5	65	26.4
11–20 points	21	8.5	44	17.9
21–30 points	9	3.7	10	4.1
31 or more points	4	1.6	3	1.2
Total	114	46.3	122	49.6

in the opposite direction. At the 100 to 109 Pintner IQ level, the Pintner IQ is the higher by 4.5 points.

In general, the greatest similarity in IQ between the Davis-Eells Games and the Pintner Test is noted in the 90 to 99 IQ range. For Pintner IQ's above 100, the Pintner IQ will tend to surpass the Davis-Eells IQ; the reverse will tend to hold true for Pintner IQ's below 80.

4) Differences in IQ as Related to Reading Level—This study was undertaken, it will be remembered, as part of a search for an instrument which would be a more satisfactory measure of the functioning of poor readers. Accordingly, the mean IQ's on the two tests obtained by pupils at varying reading levels were determined. Table 4 presents a summary of the IQ's for the 239 pupils for whom grade scores on the Metropolitan Intermediate Reading Test, Form T, were available.

It should be noted that by far the largest number of pupils are included at reading levels below 5.0. The readability of the Pintner Test is much too high for fourth and third grade readers.

TABLE 3—MEAN CHANGE FROM PINTNER TO DAVIS-EELLS IQ'S AS RELATED TO INTELLIGENCE LEVEL ON THE PINTNER TEST

Pintner IQ	N	Mean Change Pintner to Davis-Eells
120 and over	8	−28.4
110 to 119	15	−11.6
100 to 109	17	−4.5
90 to 99	57	−1.2
80 to 89	77	+2.0
70 to 79	63	+7.8
Below 70	9	+10.1

TABLE 4—MEAN IQ ON DAVIS-EELLS GAMES AND PINTNER TEST AS RELATED TO READING GRADE

Reading Grade	N	Davis-Eells	Pintner
8.0 and over	14	102.5	119.1
7.0 to 7.9	16	102.1	108.2
6.0 to 6.9	30	94.5	97.4
5.0 to 5.9	39	91.6	90.0
4.0 to 4.9	79	86.4	83.9
Below 4.0	61	81.6	77.1

At the lower reading grade levels, the Davis-Eells IQ's, which are based upon a test which requires no reading, are only slightly higher than those obtained on the Pintner Test. Even at reading grades below 4.0, the mean difference is less than five IQ points.

5) *Differences in Distributions*—It has been the experience of the Bureau that the Pintner Test, when used with the total sixth grade population in New York City, places too few pupils between IQ's 90 and 109, and too many above and below these limits, particularly too many below 90. It was felt that the Davis-Eells Games would give rise to a distribution which would more closely approximate normality. Table 5 presents a percentage frequency distribution of the IQ's obtained by the common group of 246 pupils who participated in both tests.

On the Pintner Test, approximately 30 per cent of the group tested obtained IQ's between 90 and 109, as contrasted with approximately 40 per cent on the Davis-Eells Games. The Davis-Eells Games thus overcomes the criticism of the Pintner Test that the latter places too few children in the average category.

TABLE 5—PERCENTAGE DISTRIBUTION OF IQ'S ON DAVIS-EELLS GAMES AND PINTNER TEST

IQ Level	Davis-Eells	Pintner
120 and over	1.7	2.9
110–119	5.5	6.7
100–109	13.0	5.9
90– 99	26.9	22.7
80– 89	31.1	31.9
Below 80	21.8	29.8
Total	100.0	99.9

At IQ levels below 90, both the Pintner Test and the Davis-Eells Games apparently assign too many pupils to this "below average" category. It seems unlikely that more than half the pupils are below average (below 90 IQ) in those aspects of intelligence which the Davis-Eells Games purports to measure, or that 60 per cent of the group is below 90 IQ, even in the highly verbal type of intelligence which the Pintner Test does measure.

In general, the Davis-Eells Games gives a slightly more favorable impression of the capacity of this low-average group of pupils than does the Pintner Test. The general distribution of IQ's on the two tests, however, is much the same.

SUMMARY AND CONCLUSIONS

As part of a series of studies designed to determine whether a more satisfactory measure of the intelligence of poor readers than that provided by the Pintner Intermediate Test could be obtained, the Davis-Eells Games were administered to a sample of 246 pupils drawn from nine sixth-grade classes. A large proportion of these pupils were poor readers. The major findings resulting from an analysis of the obtained data may be summarized as follows:

1) Less than a one point difference in mean IQ appears between the two measures; this difference is not statistically significant.

2) Sixty-three per cent of the pupils obtain IQ's within 10 points of each other on the two tests. In general, the greatest similarity in IQ between the Davis-Eells Games and the Pintner Test is noted in the 90–99 IQ range.

3) For Pintner IQ's below 80, the Davis-Eells IQ tends to surpass the Pintner IQ; the reverse tends to hold true for Pintner IQ's above 100.

4) The mean Davis-Eells IQ of pupils with reading grades below 4.0 is only 4.5 points higher than their mean Pintner IQ; the mean Davis-Eells IQ of pupils with reading grades between 4.0 and 4.9 is only 2.5 points higher than their mean Pintner IQ.

5) The Pintner Test assigned IQ's below 90 to 62 per cent of the group tested; the Davis-Eells Games, to 53 per cent.

The indications are that there is a high degree of resemblance between the results obtained from the use of the two instruments. When compared with the Pintner Test, the Davis-Eells Games cannot be looked upon as a markedly better measure of the intelligence of poor readers at the sixth-grade level.

ADOLESCENTS' VIEWS ON ONE ASPECT OF THEIR DEVELOPMENT[1]

RUTH STRANG

Teachers College, Columbia University

A dominant emphasis in adolescent psychology is on the adolescent's concept of himself and perception of his world. Many trends focus on this emphasis. Gestalt psychology highlights the importance of the "situation-as-perceived" by the individual. All the writings on the self-concept and "self-consistency" present an individual's view of himself as a determinant of his behavior. In client-centered counseling, the individual is helped to gain an understanding of himself and his relationships. In education, too, teachers are urged to look below the surface behavior to the meaning of the behavior of the child. From this emphasis arises the need to learn more about the inner world of adolescents and to see things through their eyes.

There are many methods of getting glimpses into how adolescents are thinking and feeling. Most widely used perhaps are the check-list type of opinion survey and the adjustment inventory. The questionnaire is a little less suggestive than the check list, but still to a considerable extent channels the individual's responses. Something may be learned from informal group discussions of adolescent problems. In such discussions and in sociodramas, adolescents have expressed some important insights into their psychology. Personal documents of various kinds are another important source of understanding. In interviews, ranging from the casual to the psychoanalytic, some adolescents will reveal a great deal of their "private world." Others, without realizing it, reveal themselves in their responses to projective techniques. The slightly structured composition combines some of the values of the personality inventory and the projective method.

There are individual differences in response to these various techniques. Some adolescents are stimulated by the group to think freely; others respond best in a face-to-face interview; still others welcome the opportunity to express their thoughts and feelings in the writing of anonymous compositions. Some adolescents are far more suggestible than others and therefore more influenced by the items on a check list or the wording of specific questions.

---

[1] Paper read at the 1954 Annual Meeting of the American Psychological Association, September 6, New York City.

In this paper, only the responses in freely written compositions will be considered. The directions for writing these compositions invite introspection on a given subject. If properly introduced they evoke frank, thoughtful, and sincere responses. Confidentialness is assured.

Questions will arise, of course, concerning the reliability and validity of the data obtained from these relatively unstructured compositions. Neither the retest nor the split-half form of reliability seems appropriate to apply to personal documents. A repetition of the request for compositions may result in showing variation within the individual rather than shortcomings of the instrument. "Very few psychologically meaningful properties ever really stay put even while the investigator is looking straight at them."[2] The split-half method obviously is not appropriate to data that are in the form of a pattern, instead of a series of discrete, comparable items.

Validity of compositions may be implied in several ways: (1) from knowledge of the rapport with the subjects in getting the data, (2) from internal evidence of frankness and lack of facetiousness, and (3) from comparison with data of a similar kind obtained from group discussions and interviews. The use of the first person ("I think", "I feel") on the basis of other evidences in the compositions seems associated with the most genuine kind of expression. The third person ("adolescents feel . . . ", "Boys and girls often think . . . ") seems associated with a measure of reluctance of the writer to give his own feelings freely. And the second person, when used, seems generally to be employed in a short story sense ("you should . . . ") by those who appear to be merely repeating lessons that they have learned from parents, teachers and other adults.

Over two thousand compositions of this kind, on a variety of topics, were obtained by teachers from adolescents of a wide range of ability and socio-economic background in grades seven through twelve. Only one of the topics—"How it Feels to Be Growing Up"

---

[2] Dalbir Bindra and Ivan H. Scheier, "The Relation Between Psychometric and Experimental Research in Psychology," *The American Psychologist*, IX (February, 1954), 70.

—obtained from 277 pupils, will be considered in this paper. Included in this sample were public high school pupils in a large city (70 in number), in a small textile town (22), in a wealthy residential suburb (19), in middle-class suburban towns (79), in a semi-rural village (14), as well as pupils in a Roman Catholic commercial high school (56) and a boys' military academy (17). The teachers presented the topic to the pupils with a minimum of structuring of the situation. The pupils wrote the compositions during the class period and did not sign their names.

There are obvious limitations to the data presented. Among these are (1) the very small number of compositions from rural schools; (2) the lack of supplementary information about the individual pupils; and (3) the differences in pupil-teacher relations which affect the quality of the responses. The first limitation may be quite easily removed in future studies; it is relatively simple and inexpensive to obtain these compositions from a large number of coöperating teachers. Every teacher who has coöperated has gained understanding of his pupils quite disproportionate to the amount of time spent in reading the papers before they were sent to the investigator. The second problem may be solved by giving each pupil a number previously identified with the pupil's name. Then the composition may be studied with reference to all the other items of information obtained about the individual pupil. The third difficulty may be handled by discarding the papers in classes where the pupil-teacher relationship was apparently poor or the standard directions for writing the composition had not been given. The papers usually give clear clues as to whether the pupil is writing frankly and sincerely or facetiously.

Two kinds of information were obtained: (1) common points of view and (2) specific psychological insights expressed by individual adolescents. The "commonality" aspect is presented in Table I, while the individuality in response is represented by the direct quotations.

The common points of view represent the thoughts and feelings uppermost in the adolescents' minds at the time of writing. To be sure, there are many other ideas that do not occur to them at the moment. Consequently, the meaning of the percentages used in this study to describe frequencies of response is not the same as in the case of check lists and questionnaires. Both present problems of interpretation. The frequency count of the compositions lacks

TABLE I—NUMBER AND PERCENTAGE OF ADOLESCENTS EXPRESSING POINTS OF VIEW ABOUT GROWING UP

Viewpoints About Growing Up	Grades					
	7–8–9 (185)		10–11–12 (92)		Total (277)	
	No.	%	No.	%	No.	%
Concern about vocation, or about the future	85	46	31	34	116	42
Feeling of increasing independence and self-direction	68	37	40	44	108	39
Awareness of increased responsibilities	53	28	42	46	94	34
Concern with social relationships	37	20	32	35	69	25
Concern with boy-girl relationships	32	17	28	30	60	22
Feelings about religion or morality	10	5	42	45	52	19
Concern with marriage and raising a family	24	13	19	21	43	16
View of growing up as fun, or "good" experience	22	12	21	23	43	16
Feeling of frustration that independence is not recognized	23	12	12	13	35	13
Awareness of increasing acceptance in the adult world	15	8	14	15	29	10
Possible reluctance to relinquish the security of dependence	8	4	16	17	24	9
Relation of sports to growing up	18	10	6	7	24	9
Problems of sibling relationships	19	10	5	5	24	9

absolute value because many individuals may hold a particular viewpoint without expressing it. On the other hand, those responding to the check list may mark a particular item with very little basis for doing so. In both instances, the nature of the sample will affect the frequency of response on separate items.

COMMON RESPONSES

Common Responses to the Topic: "How It Feels To Be Growing Up".—The tone of these compositions was generally positive and expectant. There was little evidence of the severe "strain and stress" emphasized by early writers on adolescence. Almost one sixth referred to this period as "fun" or "a nice time of life"—they had a "good" feeling about it.

Moreover, the data do not bear out the general impression of adolescent irresponsibility. They accent adolescents' desire to be independent and to take responsibility. Many of these young people thought of growing up as involving increasing independence, self-direction, and responsibility. Only a few continued to show some reluctance to relinquish the security of dependence. They are more likely to have feelings of frustration that their independence is not recognized. Some of them were aware of increasing acceptance in the adult world. The younger pupils seemed to be more optimistic and less realistic than the older group.

A surprisingly large number of junior high school pupils seemed to be concerned about vocations and the future. This high frequency of reference to vocations may be attributed to their interpretation of the topic as focused on the future instead of on the process of growing up. Another possible explanation is that the responses of some of these pupils may have been influenced by a unit on vocations recently taught in the junior high school.

The accompanying table presents, in order of frequency of mention, the categories into which the responses were classified, together with the number and per cent of responses in each. The tabulation was done by Dr. Warren Roome, who read each composition and classified each significant statement under the one heading which best described it. The headings evolved from a preliminary reading of the compositions.

The sequence of categories from junior to senior high school suggests a kind of meaningful progression in the process of growing up. As children grow older their feelings of independence and self-direction increase. They become much more concerned with religious and moral questions and with social relations, especially with boy-girl relations and marriage and family life. Their awareness of increased responsibilities grows. As they grow up physically and assume more responsibility they also become increasingly aware of their acceptance in the adult world.

Some variation was noted among the different kinds of schools in which the compositions were obtained. For example, the papers from the Catholic schools showed a high degree of uniformity. A possibly significant concept which seemed to pervade these compositions is that of sudden independence, of almost instantaneous transformation from not-adult to adult, at some specified age—either sixteen or eighteen. Possibly also of significance is the relatively high frequency with which these pupils expressed either a reluctance to leave childhood or fear of adulthood. Their religious orientation is also marked. Nevertheless, in these and even more in the other compositions unique statements of a genuinely personal viewpoint were expressed.

SPECIFIC RESPONSES

More valuable in understanding adolescents than the frequency of common responses are the psychological insights occasionally expressed by individual boys and girls. Those insights have the following claims to authenticity:

1) They are based on the first-hand experience of adolescents themselves, not on observation of their behavior by adults.

2) They depend less on delayed memory than many statements about adolescents made by adults.

3) They express attitudes and feelings directly, whereas most of the data about adolescents consists of inferences regarding their attitudes and feelings.

4) They may be culled from a large number of compositions written by teen-agers and representing a wide range of age, intelligence, socio-economic status, family backgrounds, emotional adjustment, geographical location.

A few of these insights are included here as illustrations. They are grouped under five main categories, corresponding to the more detailed categories in the table: (1) reluctance to grow up or a feeling that adults do not understand, or frustration in growing up; (2) concern about physical, emotional, social, and other aspects of growing up; (3) a somewhat neutral awareness of problems and opportunities; (4) feelings of increasing independence, self-direction, and responsibility; and (5) satisfaction in growing up and awareness of their acceptance in the adult world. The quotations have been placed under what seemed to the writer the most appropriate heading, although other points of view may also be represented in a single quotation.

1) *Reluctance to grow up, a feeling that adults do not understand or frustration in growing up.*—As shown in the table a relatively small percentage mentioned the feelings illustrated by the following quo-

tations. So far as this sampling shows, these attitudes seem to be related to the individual personality more than to the age of the pupils.

"Nothing angers me more than to have someone say, 'Oh, you are still young with no problems yet'."

"I am expected to have manners like an adult, yet have to go to bed early like a little child."

"Parents say, 'You're old enough to help out with this,' and 'You can't do that; you're much too young'."

"More often you wonder what it will be like when there is no mother and father to guide you and it scares you."

"Growing up is a hard thing to face because childhood is so dear to our hearts."

"The feeling that comes with growing up sometimes frightens you. The thought that now you are responsible for all your actions. People can't say, 'She is only a child and you can't blame her for her bad actions'."

"My friends and I have, at one time or another, felt that our parents did not understand us, nor care to."

"Grown-ups say they remember how it was to be growing up: that they understand our problems. I disagree, as they forget the unpleasant memories; only someone who is going through the process knows what it is like to be growing up."

"When I finish growing up, which I doubt anyone really does completely, I hope I will be able to remember how I felt so that I can understand people who are growing up then."

2) *Concern about physical, emotional, social, and other aspects of growing up.*—In these quotations, which are typical of a large percentage of the compositions studied, points of views presented by psychologists concerning uneven growth and acne, boy-girl relations, swings of mood, and alternation between childish and adult interests are reinforced.

"The age of thirteen is a difficult age . . . . When boys are thirteen they are too old for cowboys and Indians and too young for girls. Girls are too old for dolls and too young for boys."

"Such disadvantages as legs too long, arms too long, pimples, and many other things that make a boy feel very out of place."

"Growing up has made me feel different about religion. When I try to understand and ask questions, I either get a confusing answer or my question cannot be answered."

"Growing up in the world today means an uncertain future. It means bringing our own children into a troubled world. Waiting years for boys who may never return from war results in emotional upset for many."

"One of the difficulties of teen-agers is their moods. Sometimes they get up in the morning feeling wonderful. Instead of talking, they feel like singing. Other days they feel wretched and depressed. They can hardly drag themselves around . . . . These moods come and go in a mysterious way, for just no reason at all."

3) *A somewhat neutral awareness of problems and opportunities.*—Instead of concern about their problems, some of the youngsters take an understanding, accepting, and even humorous attitude toward them. This insight into the growing-up process is illustrated by the following quotations:

"Parents, it seems to me, worry too much about our welfare. But that is simply because they are under a false impression; we are grown up long before they have a chance to realize it."

"Most girls think about only one thing—boys."

"Every time a boy pays a nice compliment, I feel that I've grown up more. It makes me think that in some way I've learned to better my appearance."

"Some girls when they are quite young are regular tomboys. Then all of a sudden, something happens. They become conscious of their appearance. They start to take care of their hair . . . they start to hang up their clothes . . . they begin to notice different styles. They wear their first pair of high heels and start to apply lipstick."

"My brother is three years older than I am . . . . Wherever I go I no longer am Margaret Kerr but Sam's sister. My brother and I get along swell. Even though he calls me 'Stinkweed' I know it's just his peculiar way of showing me affection."

4) *Feeling of increasing independence, self-direction, and responsibility.* Many illustrations of this point of view might be quoted. In fact, in this sampling of adolescents it was the dominant way of perceiving the process of growing up. They expressed simultaneously their desire to be free from parental domination and restriction of childhood and their eagerness for more mature experiences.

"When a boy reaches his teens he doesn't like to be led around by his mother's apronstrings and he wants freedom and likes to feel self-dependent."

"Instead of running from mistakes as a child would, I now see the wisdom of facing and conquering error."

"I like the idea of being able to choose and make decisions for myself."

"As I grow older I can do more things without my parents having to worry about me."

"Growing up means . . . facing problems with almost no help."

"My problems are now solved by me. I no longer run to Mom to solve them unless her advice is absolutely needed."

"As you grow up you are not so much of a burden to your parents, and they do not worry about you but pass on the worries to you and you carry your worries for yourself."

"When a problem arises, big or small, you can work it out without getting yourself and everyone around you confused and nervous."

"I am beginning to understand the responsibilities of my position, and most amazing of all, to tackle them."

5) *Satisfaction in growing up and awareness of their acceptance in the adult world.* Although their desire for independence is much more frequently mentioned than their satisfaction in the process of growing up, almost one sixth referred to the rewarding aspects of adolescence and one tenth recognized their increased acceptance by adults. The following quotations refer especially to the ego satisfaction of being able to handle the problems of adolescence and being accepted on a mature level. In much of the writing on adolescence these positive feelings about growing up are overshadowed by the emphasis on anxiety, insecurity, and "problems."

"Growing up makes me feel as if I were somebody important."

"I derive a great satisfaction from the solution of my problems and even anticipate the advent of new ones with a curious mixture of anxiety and exhilaration."

"As you reach your teens you find that people will talk to you and let you give your opinions."

"I am now on eye level with adults and beginning to be accepted as a person in my own right."

The wide range of attitudes toward growing up is represented by these two responses: (1) "I think that growing up is one of the most exciting things that can ever happen to anyone," and (2) "On the whole growing up is the biggest ordeal anyone goes through. I believe I have had more trouble now than I ever had. Disappointments and let-downs are the biggest hurdle in growing up. One thing I learned to do is never build up my hopes."

The statements made by adolescents about themselves are appealing; they are directly, simply, and often vividly stated. Many of them verify the viewpoints of psychologists about adolescents. Some of them suggest positive feelings and attitudes that perhaps have not been sufficiently recognized in adolescent literature—(1) the satisfaction that comes from successful accomplishment of adolescent tasks, (2) the recognition that responsibility should accompany freedom, (3) their eagerness to assume responsibility, and (4) their insight into adults' points of view. Most of all, these compositions alert us to individual differences among adolescents and the diversity of viewpoints and feelings which they express.

EFFECTS OF VARIATIONS IN TEACHER RÔLE BEHAVIOR ON THE STUDENT-TEACHER RELATIONSHIP

MARTIN M. GROSSACK

Philander Smith College, Little Rock

The concept of social rôle has become one of the most widely used explanatory tools in social science. Parsons *et al.* have theoretically described several related dimensions of social roles.[1] In western society, rôle expectations differ significantly in the areas of formal organizational behavior in comparison with the demands of kinship and friendship situations. Large organizations, such as educational systems, theoretically require behavior that is universalistic, non-emotional and specific in function. On the other hand, situations of kinship and friendship carry with them rôle expectations of particularistic orientation, emotional relationships and diffuseness of function.

An example will illustrate the differences inherent in the two types of situations. The teacher is expected to treat all students alike regardless of whom they are (universalism), show no emotional involvement in their conditions or towards them as individuals (there is, for example, a general taboo on student-faculty dating), and his authority is specifically limited to problems of teaching (he has no power over student's religious or private behaviors as long as they do not affect the behavior of the student *as a student*).

The experiment reported here is concerned with the applicability of this conceptualization to the student-teacher relationship. Three predictions were derived for empirical test:

1) Emotional ties will develop more frequently in conditions of particularism than in those of universalism.

2) Respect of status differences will be less frequent in conditions of particularism than in those of universalism.

3) Barriers to non-relevant communications will be greater in conditions of universalism than in those of particularism.

---

[1] T. Parsons and E. A. Shils, Eds. *Towards A General Theory of Action.* Cambridge, Massachusetts: Harvard University Press, 1951.

PROCEDURE

Eighty-one Negro Ss were made available to E during a meeting of students. Ss were randomly given one of the following mimeographed characterizations:

A. *Particularistic Form.*—Professor Jones was brought up in a small Midwestern town. Then he attended a small college. As a teacher he tends to form close friendships with his students and gets emotionally involved in his teaching and his relations with students.

It is known that Professor Jones takes part in many social activities with students and calls them by their first names. He shows some preferences in his relationship with students and possibly treats a few better than he does the rest.

B. *Universalistic Form.*—Professor Jones was brought up in a big city and attended a large Midwestern University. As a teacher, he is very careful to keep his distance from any close friendships with students. He just teaches his courses without emotional involvement in relations with his students. Outside of class, he has as little as possible to do with his students. He makes it a practice to treat all students alike and calls them Mr. or Miss, rather than by their first names.

This procedure, in effect, presented two ideal-types and assumed there would be differences in response to them. Each S received the same fourteen-item questionnaire.

RESULTS

The differences in questionnaire responses by recipients of the particularistic and universalistic teacher characterizations are shown in the table.

1) *Emotional ties hypothesis.*—The prediction that emotional ties will develop more frequently in conditions of particularism than in those of universalism received some support from the results in Table 1. Ss were more likely to feel friendly towards the particularistic characterization (P.10) and slightly, but not significantly, more likely to answer items 2, 4, 8, 11, and 14 favoring it.

2) *Respect of status hypothesis.*—The prediction that respect of status differences will be less frequent in conditions of particularism than in those of universalism was supported. Ss were more likely to respect the universalistic characterization and feel he grades fairly (P.10) and slightly less likely to consider him a "small guy".

TABLE I—DIFFERENCES IN QUESTIONNAIRE RESPONSES CONCERNING TEACHERS WITH PERCEIVED PARTICULARISTIC AND UNIVERSALISTIC RÔLE BEHAVIOR

Question	Characterization				χ^2	P
	Particularistic		Universalistic			
	Likely	Not likely	Likely	Not likely		
How likely would you be to:						
1. Tell Professor Jones your personal problems?	29	17	13	20	3.30	.10
2. Attempt to develop a friendship with him?	37	8	23	10	*	*
3. Joke with him?	32	12	9	24	13.50	.001
4. Take a course with him?	34	12	19	13	*	*
5. Think he is a small guy?	25	20	12	19	*	*
6. Think he deals smoothly with people?	30	15	22	11	*	*
7. Tell him your disagreements with him?	39	8	21	12	2.90	.10
8. Think he has a healthy personality?	33	14	19	13	*	*
9. Feel friendly towards him?	41	4	20	8	3.13	.10
10. Think he grades students fairly?	27	18	27	6	3.31	.10
11. Think he likes the students?	39	7	26	7	*	*
12. Learn a lot in his class?	29	16	22	10	*	*
13. Respect him?	39	7	32	0	3.63	.10
14. Feel he knows his stuff?	33	12	25	11	*	*

* Not significant.

3) *Barriers to communication hypothesis.*—The results suggest that universalistic rôle behavior will serve as a barrier to non-relevant communications. Ss would be less likely to joke with the universalistic professor (P.001) or to tell him their personal problems (P.10). However, the formality of the rôle relationship might also serve to minimize the amount of disagreements that are expressed overtly to the universalistic authority figure (P.10), suggesting that the particularistic individual might less likely have people griping without his knowledge.

DISCUSSION OF FINDINGS

The implications of these results for educational psychology merit consideration. Although the statistical evidence is suggestive rather than conclusive, the social psychological factors in the student-teacher relationship have received scant empirical attention from educators; therefore, any evidence merits attention.

First, it is important to raise the question of the inherent incompatibility of the two types of rôle orientations. Do students having close friendships with teachers believe that they receive the same impartiality in grading as non-friends? Theoretically, such impartiality is not possible for friendship relations. Can the teacher who judges students' performance really give emotional support to all class members? It would seem that the rôle of teacher *as teacher* should be limited to non-therapeutic activities.

The results show that students will more readily develop positive affect towards particularistic teachers and more readily confide in them. Impersonal teachers, that is, those who deliberately take a universalistic orientation, have to realize the necessity for creating a congenial working climate to offset the formality of their rôle behavior. Possibly, it is best to encourage the students to feel friendly and confide, but good sense dictates that the teacher should not reciprocate to the extent of an emotional involvement that could disrupt his rôle behavior as teacher.

Rôle-playing should probably receive more attention in teacher training programs so that some of the conflicts inherent in the rôles of teacher can be anticipated by future teachers. A body of reliable knowledge is necessary and urgent if our knowledge of how to teach and to relate to students can be based on evidence rather than tradition.

SUMMARY

From Parsons' theoretical description of particularistic and universalistic dimensions of rôle behavior, two ideal characterizations of college professors, one exemplifying the former orientation, the other the latter, were presented to student Ss. Differences were found in the perceptions and in the expected communications students would make to professors of each type. Evidence was presented supporting hypotheses of fewer emotional ties, more respect of status differences, and more barriers to communication towards the universalistic characterization.

TEACHER REACTIONS TO CONFLICT SITUATIONS

SAMUEL I. SPECTOR

Center for Human Relations Studies, New York University

In the practice of human relations, since every individual is motivated by his inherent and acquired frame of reference, personal and therefore varied responses are made to given stimuli. In the general scope of interrelationships in which each individual does not have continued daily concentrated contact with one or more fellow human beings, the impress of his frame of reference may be transitory. In other words, a mood or attitude may be evoked during the contact, but it is likely to disappear when the stimulus is removed. However, when there are daily meetings as in teaching, continual exposure to human relations conflicts, especially as related to supervisor, teacher, and pupil interaction, presumably leaves a more permanent impression. Here, success in human relations cannot be made dependent on the hope that somehow they will muddle through. Since the full development of children is involved, there is too much at stake. And, further, since in this development, the teacher is the focal point, his ability to interact successfully becomes a matter of prime importance. With this thought in mind, the study described below was initiated.

At the outset these questions arose: What conflict situations would be common to the majority of teachers? What would be the measure of behavior responses? Would teachers react uniformly or diversely to a given conflict situation? Would they show a lag between their concept of what was the right thing to do, what was the practical thing to do at the moment of conflict, and what they actually did? What evidences of authoritarian behavior would be apparent? What are the implications of the uniformity or diversity of response for teacher training?

It was felt that a questionnaire might be the suitable instrument, since there was no intention to probe deeply into individual characteristics. The purpose of the study was to determine how a number of teachers would react to the same stimuli. On the basis of talks with supervisors, teachers and pupils, conflict situations were determined. There were five for each relationship (teacher-supervisor, teacher-teacher, and teacher-pupil), classified as to origin (self initiated, initiated by other individuals, and initiated by circumstances). The subjects were asked to state what they would do, should do, or did in terms of suggested courses of action listed by letter and number, in the situations outlined. Thus in the teacher-supervisor relationships, the following were posited:

1) Your principal is interested in a certain philanthropy. He wants all teachers to subscribe. You do not care to participate. He presses you daily.

2) You have invited your supervisor to observe you. He visits for an hour and leaves behind a very uncomplimentary report. He refuses to discuss the write-up with you.

3) The supervisor is riding a weak teacher. You cannot tolerate this treatment and intervene with the supervisor. He tells you to mind your business and plainly shows his resentment at your interference.

4) The supervisor suggests a certain teaching procedure. You know that it will not work in your class. He claims superior knowledge and insists you use his suggestion nevertheless. Tempers flare.

5) As a disciplinary measure, an older child from a higher grade has been placed into your class. He gets into a quarrel with one of your pupils and injures him severely. The principal holds you completely responsible for the difficulty.

Numbers 1 and 2 represent situations initiated by other individuals, numbers 3 and 4, those that were self initiated, and number 5 a situation arising from circumstances.

In the teacher-teacher category, the following situations were suggested:

1) You have been accused of spreading gossip about another teacher. You are entirely innocent. The teacher confronts you and berates you.

2) The teacher in the adjoining room is out of her room often during the day. At these times, the noise from her class disturbs your work. You remonstrate with the teacher. She tells you to stop being a goody-goody and kill-joy.

3) A coarse and shallow fellow teacher, whom you do not like, seeks your friendship. He goes out of his way to meet you with his attentions and small gifts. Despite all your efforts to discourage him he keeps on trying.

4) There is a coveted assignment that you want, but that should go to a fellow teacher because of his seniority. You pull strings and get the assignment. Whenever he sees you he shows his resentment by his looks and remarks.

5) You have been away from school for two weeks because of illness. The school Social Committee, which usually recognizes such absences by a gift, has failed to do so in your case. Upon your return, the committee lamely and weakly mentions the oversight.

Numbers 1 and 3 represented initiation by other individuals, 2 and 4 initiation by the teacher, and number 5 initiation by circumstances.

The third group of situations concerned teacher-pupil interactions. They were as follows:

1) You are irritable. A child drops a book accidentally. You berate her. Stung by your injustice, the child glowers and mutters under her breath.

2) While the supervisor is present, the smart-alec in your class decides to show off. In answer to your every gesture and silent plea, he shows off still more. He mimics you.

3) One of your pupils comes to school in ragged clothing. You obtain some good second-hand clothing, and, in privacy, offer him better clothing. The child bursts into bitter resentful tears and runs out of the room.

4) One of your pupils seems to show an instinctive dislike for you. He goes out of his way to irritate and annoy you. His special pleasure seems to be to mispronounce your name purposely.

5) You are on duty in the yard. Some large boys and girls from the graduating class are fooling around. They do not see you approach. When you scold them they get fresh, not realizing that you are a teacher. The boys, to save face, try to act tough with you.

Here, too, the genesis is self initiation for numbers 1 and 3, other initiation for numbers 2 and 4, and circumstance initiation for number 5. It should be noted that each of the fifteen situations was a part of the writer's personal or vicarious experience.

The formulation of a continuum of courses of action was the result of many castings and recastings in the effort to graduate behavior patterns ranging from the immature physical exchange of blows to the objective attempt to determine the cause of irritation. Some degree of continuum was obtained, but exclusiveness of the various gradations was difficult to achieve. The courses of actions are as follows:

a.1. Exchange blows	2. Threaten with fists	3. Curse vocally
b.1. Talk back	2. Bluster	3. Threaten 4. Curse under the breath
c.1. Laugh it off	2. Placate by smiling	3. Be coldly polite 4. Use soft words 5. Divert attention to something else
d.1. Remain calm and silent	2. Ignore	3. Walk away
e.1. Feel annoyed	2. Feel resentful	3. Feel hurt
f.1. Discuss calmly	2. Stick to the facts of the situation	3. Try to determine the cause of the irritation

The fact, that except for a.1 and a.2, all courses of action were selected by one or more subjects, seems to give some validity to the continuum.

The format of the questionnaire deserves some description. It consisted of two pages, each headed by the list of courses of action. At the very outset, the purpose of the questionnaire was stated in these words:

> "The purpose of this questionnaire is to determine how teachers feel they should and would act and how they did act in real or possible situations. Alongside each situation is placed a form containing three blank lines for the three possibilities—you should do; you would do; you did. From the list of possible courses of action listed below, you are asked to select the letter and number of the action you would choose to suit the situation; and to place them in the respective blank lines representing the possibilities."

On the left side were set three columns of possibilities, and on the right the five situations. The example below is indicative.

Possibilities			
You should do	You would do	You did	*Situation*
___	___	___	1. Your principal is interested in a certain philanthropy. He wants all teachers to subscribe. You do not care to participate. He presses you daily.
___	___	___	2. You have invited your supervisor to observe you. He visits for an hour and leaves behind a very uncomplimentary report. He refuses to discuss the write-up with you.

The same format was repeated on the second sheet for teacher-

teacher and teacher-pupil situations. At the bottom of this sheet, information was requested about the age of the subject (above or below 40), the number of years' teaching experience and the number of schools in which the subject had taught. Age was asked to determine whether a relationship existed between this factor and the course of actions selected. The same rationale motivated the question on teaching experience. Finally, there was the possibility that teaching in a great number of schools indicated an instability, which would be represented in choices of more immature courses of action. Signatures of the teachers were not solicited so that free and uninhibited choices might be made.

The questionnaire was administered to fifty-three teachers from five different schools, servicing the same type of lower middle-class population. There was no attempt to differentiate between the sexes. Of these teachers, eleven indicated that they were below forty years of age, and twenty three that they were above that age. The average number of years of teaching experience was eighteen. The average number of schools taught in was three. Careful examination of the choices made did not indicate a positive relationship between age, experience, or number of schools and the selection of courses of action.

In terms of situations, it appears that the greatest number of choices for any of them did not exceed fifty per cent. This means that, for the most part, fewer than one half of the teachers made the same choice of a course of action—a great diversity indeed. This variety of response is indicated by the following breakdown, in terms of the most selected behavior reaction f.1 or "discuss calmly".

TEACHER-SUPERVISOR RELATIONSHIPS

	Should Do	Would Do	Did
Other-initiated situations	28%	19%	20%
Self-initiated situations	40%	20%	20%
Circumstance-initiated situations	50%	40%	35%

(READ: 28% of all teachers in other-initiated situations chose the f.1 category, etc.)

At this point, the decrease between the "should do" and "would do" possibilities should be noted. This seems to indicate a lag between values (should do) and the application of these values (would do). Of interest, too, is the greater claimed (by the subjects) ability to apply calm discussion to circumstances—initiated conflict situations. At least half of the teachers could not agree upon one course of action.

TEACHER-TEACHER RELATIONSHIPS

	Should Do	Would Do	Did
Other-initiated situations	22%	24%	30%
Self-initiated situations	31%	20%	33%
Circumstance-initiated situations	12%	1%	1%

(READ: 22% of all teachers in other-initiated situations chose the f.1 category, etc.)

Here again a lag between values and their application is evident, though to a much lesser degree. It is especially apparent in self-initiated situations. At least two-thirds of the teachers could not agree upon one course of action.

TEACHER-PUPIL RELATIONSHIPS

	Should Do	Would Do	Did
Other-initiated situations	13%	12%	20%
Self-initiated situations	25%	20%	25%
Circumstance-initiated situations	27%	21%	31%

(READ: 13% of all teachers in other-initiated situations chose the f.1 category, etc.)

The lag between should-do and would-do choices continues in this category also. Again at least 69% of the teachers could not agree upon one course of action. Along with Teacher-Supervisor Relationships it shows the greatest choice agreement in circumstance-initiated situations.

From another point of view, these statistics are of interest in illustrating diversity of responses. Of fifteen possible responses in the Teacher-Supervisor category, 9 or 60% were indicated for behavior pattern f.1 (discuss calmly), 3 or 20% for f.2 (stick to facts), while the rest were scattered. In the Teacher-Teacher relations, of the 15 possible responses 7 or 46% were indicated for f.1 (discuss calmly), 3 or 20% for c.2 (placate by smiling), while the rest were scattered. In the Teacher-Pupil interactions, of 15 possible responses 5 or 34% were indicated for f.1 (discuss calmly), 5 or 34% were indicated for d.2 (ignore) and the rest were scattered. It is interesting to note the progression in the f.1 choices—34% for Teacher-Pupil relationships, 46% for Teacher-Teacher relationships and 60% for Teacher-Supervisor relationships. It is heartening to note that so many teachers feel that they can discuss conflict situations calmly with their supervisors.

Another breakdown in terms of greatest choices indicates the following choices of courses of action:

COURSES OF ACTION

Teacher-Supervisor	b.1						d.1	d.2	e.1	e.2	e.3	f.1	f.2	f.3
Teacher-Teacher		c.1		c.3			d.1	d.2	e.1			f.1	f.2	f.3
Teacher-Pupil			c.2		c.4	c.5	d.1	d.2	e.1			f.1		f.3

Here the majority indicated the following reactions: d.1 (remain calm and silent); d.2 (ignore); e.1 (feel annoyed); f.1 (discuss calmly) and f.3 (determine cause of irritation). In Teacher-Pupil relations only 2 of 15 possible f.3's were indicated. If this reaction is most desired, the teachers are sadly wanting. The emphasis on f.1 (discuss calmly) does not presume constructive discussion. The choices of c.4 (using soft words) and c.5 (diverting attention) in Teacher-Pupil relationships is of interest, but they represent less than 27 % of the total responses. This is also true for the c.3 (being coldly polite) choice in Teacher-Teacher relations. Here for Situation 3, as much as 43 % of the choices were involved.

In terms of what the teachers thought they should do and what they would do, there are interesting comparisons. In Teacher-Supervisor relations, 42 % of the teachers said that they should discuss Situation 1 calmly, but only one half of this number felt they would do so. Approximately the same response was made for Situation 4. Fifty per cent felt that they should and would discuss calmly Situation 5. Further in Teacher-Teacher relations, 42 % declared that they should discuss calmly Situation 1, while 47 % felt that they would do so. For Situation 2, the comparable percentages were 42 and 30. For Situation 3, about 40 % said they should be coldly polite, while 50 % said they would be so. Finally in Teacher-Pupil relationships, 25 % of those answering said they should and would ignore in Situation 1. This was also true for Situation 5. In Situation 3, 34 % said they should discuss calmly, while 29 % said that they would. In general, it would seem that in all categories, Situations 1, 3 and 5 received the greatest number of choices. All of the foregoing attests to the diversity of choice of behavior patterns.

On the premise that individuals who made "a" and "b" action choices (exchange blows to curse under breath) were immature and authoritarian personalities, fifteen questionnaires containing such responses were analyzed in detail. It was determined that of these 15, 6 were over forty years old and 3 under this age. This compares with 23 and 11 for the total population. Six did not give their age, a possible indication of suspiciousness. The average teaching experience of 10 of the 15 (5 did not answer) was 18 years, the same as for the total population. The proportion for number of schools taught in also remained constant. It is thus evident that in the study, age experience and number of schools taught in were not affective factors in the authoritarianism of the personalities. In the range of choices for all categories (Teacher-Supervisor, Teacher-Teacher, Teacher-Pupil), a.1 (exchange blows) and a.2 (threaten with fists) choices were absent. All other behavior reactions were included. The a.3 action (curse verbally) was indicated as a "would" choice by one teacher for Situation 4 in Teacher-Pupil relationships and by another teacher for a "would" choice for Situation 2 and a "should" choice for Situation 4 in Teacher-Supervisor relationships.

Authoritarianism showed itself most in Teacher-Pupil interactions when 13 "did" responses, chiefly in the "threaten" behavior pattern, were indicated. It appeared less in Teacher-Supervisor relationships, again with no duplications, and not at all in Teacher-Teacher relations. In the Teacher-Supervisor category, the chief reaction chosen was "talk back". The same was true for the Teacher-Teacher category. In the Teacher-Pupil category, threats and talking back were made chiefly for Situations 2 and 5; in the Teacher-Teacher category talking back was indicated for Situation 1; while in Teacher-Supervisor relations, talking back was chosen for Situations 4 and 5. In all three categories, the teachers were able to tell most readily what they would do as indicated by a 91% response, less readily what they should do as indicated by a 75% response, and least readily what they did as indicated by a 63% response. The teachers showed the greatest frustration in #5 Situations, those that were caused by circumstances beyond their control. In the Teacher-Pupil category, the great choice of Situation 2 which involves the presence of a supervisor is interesting for its interactive possibilities, i.e., fear of or discomfort with the presence of the supervisor made the security of the teacher less stable and helped project this feeling into the reaction toward the pupil. The large choice of Situation 4 in the Teacher-Supervisor category, which involves a self-initiated dispute about the supervisor's asserted superior knowledge of teaching procedure, indicated authoritarian personality traits in that the teachers felt that their knowledge and status were under attack. Many of the teachers showed immature behavior when accused wrongly by a fellow teacher. Fifteen out of 53 cases, a percentage of 30, showed evidence of immature behavior. Finally, it appeared that the greatest number of choices over the widest gamut of choices indicative of authoritarian behavior patterns, was to be found most in Teacher-Pupil relations, less in Teacher-Supervisor interactions and least in Teacher-Teacher relationships.

What, now, are the implications of the study, especially for teacher training? In a pilot study of this kind, based on so small a sampling, no definitiveness is possible. However, certain trends are apparent. If the teachers studied are representative, and more than half of them did not react in the same manner to a variety of conflict situations, what would be the effect on the emotional growth of our children? Further, to what extent will this disparity in behavior reaction between teacher and supervisor and fellow teacher in turn affect teacher-pupil relationships? Is the lag between what the teachers felt they should do and what they indicated they would do, an indication of the failure of the teacher-training program? Is it the business of the college to develop in embryo teachers an emotional objectivity that would make it natural for them to try to determine the cause of irritation in cases of conflict? These questions merely scratch the surface of possible implications. However, what does arise from the study is the need for teaching concrete human relations courses in the various teacher training institutions. These would involve the resolution of major individual emotional problems or weaknesses through mental hygiene courses, group therapy or even individual psychiatric attention. They would include the establishment of clear, cogent values in education and democracy, made so vital, that the lag between their conception and application would be reduced to a minimum. Finally, they would provide for the analyses and treatment of concrete human-relations conflict situations in all phases of the teacher training, from the philosophy of education course to student teaching. This is the meat upon which good teacher tyros must feed!

Book Reviews

Doris M. Odlum. *Psychology, The Nurse and the Patient.* New York: Philosophical Library, 1952, pp. 168.

How psychology has helped and can help nursing in England is what the book is about. It is a revised edition of a syllabus used for preliminary and final examinations for State registered nurses in London. The author has played the rôle of physician, psychiatrist, lecturer to nurses and also of patient, and thinks she has gained enough insight from all of this to transmit to others. The material is presented in nineteen short chapters. The more practical ones which will more likely make sense to the American reader are the chapters on the Patient, the Hospital and the Patient, the Nursing of Children, Treatment of Psycho-Neuroses. The theoretical considerations of the kind of psychology that can be helpful to nursing will be puzzling to some American psychologists and questionable to others. The chapter on the How and Why of Human Behaviour and Inborn Mental Factors more or less accepts the McDougall list of instincts and supplements it with other more generalized lists. Here are given as ways of helping nurses a list of instincts and emotions: flight with fear, repulsion with disgust, sex with tenderness, etc. Many facts are presented here as if the last answer was in, although there are still plenty of question marks on them. Maybe it's advisable for a nurse to just do what is being done or told to do rather than ask many questions about the limitations of some of the treatments listed here. For general purposes, there is no doubt that this volume has served and can serve English nurses in training. For more reliable knowledge of what psychology has to teach about human behaviour that could be of use to nurses, another book can be written.

H. Meltzer

Psychological Service Center, St. Louis, Mo.

The Kentucky Symposium. *Learning Theory, Personality Theory, and Clinical Research.* New York: John Wiley & Sons, 1954, pp. 164.

At the initiative of the Psychology Department of the University of Kentucky and with the support of the University administration, a very worthy effort was made to bring together a number of psychologists for the purpose of clarifying and bringing integration into the points of view and the theories relating to the three areas indicated in the title. The Kentucky Symposium was conducted and later found publication in this volume under the direction of Dr. Frank A. Pattie.

The subjects with their authors are as follows in their respective order: Current Interpretations of Learning Data and Some Recent Developments in Stimulus-Response Theory (Kenneth W. Spence); Stimulus-Response Theory as Applied to Perception (Selos D. Wickens); Motivational Forces Underlying Learning (Harry F. Harlow); The Premature Crystallization of Learning Theory (Norman R. F. Maier); Learning and Explanation (Donald K. Adams); Ego Psychology, Cybernetics, and Learning Theory (O. H. Mowrer); Personality Structures as Learning and Motivation Patterns—A Theme for the Integration of Methodologies (Raymond B. Cattell); Prospects and Perspectives in Psychotherapeutic Theory and Research (John M. Butler); Learning: an Aspect of Personality Development (Donald Snygg); Errors: Theory and Measurement (R. B. Ammons), etc. This kind of collection may be of great value to psychology. It should help very much in the development out of the 'psychologies' into the psychology which is not yet quite in the thinking of many workers in psychology who are still too limited and too sectarian.

Careful reading of this symposium may emphasize the fact that some theorizing has been premature, that much more experimental and clinical data are needed, but, on the other hand, that theoretical considerations have not only been necessary for proper investigation, and that more comprehensive theories are necessary for the further development of psychology.

The study has the advantages and disadvantages of multiple authorship, but the former outweighs the latter; the reader can better evaluate the present situation after considering the critical comments introduced by the writers. There is perhaps as much agreement as could be expected in the present development of the science. We are having the difficulties of terminology well known to the other sciences. We are dealing with exceedingly complex problems. We have probably succeeded in thinking up more points of view and modes of approach than any other science. And we have certainly not failed to fall back on the resources of the English language in such ways as have seemed to many too complicated and even at times pedantic. Considering the complexity and difficulty of the task undertaken, this volume can be thought of as clear, direct, not too technical, and not pedantic.

The special recommendation for the book, is, the reviewer thinks, in the success it has achieved in the attempt to relate theoretically several fields or areas of psychology, and the stimulus it should give to better thinking and investigation in terms of a wider, more comprehensive view. With this kind of progress in psychology, no one need despair of the science except he who has hedged himself in with too narrow and puristic a definition, and has found that psychology has already burst the bounds of his definition and gone on like a living organism to become what was not dreamed of in the ancient armchair.

The reader will find very frank, honest, sometimes devastating, but also constructive, evaluation of theory and theorizing in psychology. Some psychologists at least are willing to bring out the shortcomings of psychology and of psychological workers (shall we say especially of other psychologists?) but are as eager to point out the possible ways of improvement as far as they are understood.

As these authors have disagreed with one another, so the reader can find his particular points of disagreement, but as a total contribution, he should find an excellent critical evaluation of many psychological matters that will well repay his careful study.

A useful Index (pp. 161–164) concludes the book.

A. S. Edwards

The University of Georgia

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume 46 December, 1955 Number 8

GROUPNESS OF THE GROUP[1]

IRVING LORGE

Teachers College, Columbia University

How an individual makes his choice of a vocation or a profession is of great concern to educational psychologists; how an educational psychologist makes that choice should be of significance to all of you. I would like to bring to your consideration the way Dr. Edward Lee Thorndike became an educational psychologist. On completing his doctorate on animal intelligence at Columbia University, he states in his autobiography,

"In the Spring of 1898, I was offered two positions, one as a teacher of psychology in a normal school, the other at a much lower salary as a teacher of education in the College for Women at Western Reserve University. I chose the latter, partly because my brother expected to go there and partly because of the repute of Western Reserve. I spent the summer in reading the facts and important theories about education and teaching. This could then be done if the history of educational practices was omitted. After a year at Cleveland, I was given a trial at teaching psychology and child study at Teachers College, and there I have spent the past thirty-one years."

It would be supererogation, indeed, to tell you that Dr. Thorndike, as an educational psychologist, made contributions to child study, mental and social measurement, the psychology of school subjects, mental inheritance, individual and sex differences, the organization of abilities, memory, work, interest, selection and classification of personnel, adult learning, human nature and the social order, language and communication. But I have supererogated! If Dr. Thorndike had been a joiner, it is reasonable to assume that he could have met the requirements for fellowship in practically all of the divisions within the present structure of the American Psychological Association.

---

[1] Presidential address read at the meeting of Division 15 of the American Psychological Association in September 1954.

But if he had, there would have been a high likelihood that he would not have been quite as productive as he was. Perhaps contemporary psychologists are chary of multiple membership in the association because their productivity, too, may be impaired. Certainly, if one accepts the factorial analysis of the membership of the APA reported by Dr. Dorothy Adkins, we have the evidence that educational psychology has a strong association with childhood and adolescence. As she put it, "Factor 7 has high loadings only on childhood and adolescence and on educational psychology, and neither of these has a high loading on any other factor. . . . Factorial results definitely argue for the union of childhood and adolescence only, perhaps under the title, 'developmental psychology'."

There can be little doubt of the considerable overlap in the affiliations of the members of Divisions 5 and 15. Yet, it must be re-emphasized, that affiliation and interest are not yet perfectly correlated either within the association's membership or in the population itself. Structure recommended on the basis of affiliation not only tends to maintain *status quo* but also may defeat the original purpose of the several divisions. Certainly, one would expect that new fields in psychology must be pioneered and developed by zealous psychologists banded together in a common purpose of improving knowledge and its applications.

Some psychologists have suggested the formation of a smaller number of sections because of the ease of programming. Because of common affiliations, all papers and symposia related to the common membership, they say, could then be scheduled together. This, to me, seems a dangerous principle. The separateness of the divisions does suggest that programming be on a successive rather than on a concurrent basis. Indeed, the overlap of educational psychology on all other divisions is so great that it can be said that for educational psychologists, nothing in psychology is alien to them. Educational psychologists, as a group, are interested in the content, methods, knowledges, and applications of each division to the improvement of teaching and of learning. Affiliations, assuredly, do not define the group.

My reference to Dr. Thorndike's choice of career and productivity in his chosen field was not only evidence of my intellectual indebtedness to him, but was intended to show that from the educational psychologists there emerged the leadership for the development of many of the currently strong divisions in our association. Educational Psychology still covers the wide range of the applications of psychology to the improvement of teaching and learning from childhood to maturity and in settings varying from informal teaching and learning within the family through the more usual and more structured school situations for the transmission of the cultural heritage to pupils and students, to the tremendous variety of informal experiences by which individuals adjust to their physical and social environment.

Indeed, Thorndike, in his work as an educational psychologist, tended to accept the implied objectives and stated content of schools as the given. Given the objectives, he tried to estimate the ways in which psychology could help teachers to bring learners to realize the curricular goals. His researches demonstrated the weakness of some objectives, the error in the rationale for some methods, and the inefficiency and unreality of much of the content and skill taught children in school in the early decades of this century.

He suggested the manner for introducing reality in subject matter, usefulness for skills, and science in methodology. His procedure was that of the researcher who investigated to find out if it is so, does it work, will it last? There is a tendency, among educational psychologists, today to accept some objectives, procedures and content as legitimate curricular goals without assaying the attainability of them. For instance, there seems to be considerable emphasis in the educational and psychological literature on the attainment of groupness and group dynamics. There can be little quarrel with the stipulation of such goals provided that some attempts be made to find out whether the goals as ends, or as means to other ends, are realized or realizable.

From a brief review of group dynamics, it is evident that much attention has been given to the purpose of group dynamics at the elementary school level. The expressed purposes range from the mere learning of specific skills and knowledges to the attainment of the broader aspects of democratic living. For instance, Thelen believes that instructional efficiency can be increased through the teaching of group dynamics in the classroom. He suggests that all students in a subgroup can be brought up to the level originally possessed by the most skilled person in it, if group methods of teaching are intelligently used. It is true that Thelen feels that a primary purpose of group methods in teaching and learning is in the promotion of creative thinking and social organization. When the class discusses and settles the question of how work shall be organized and individual efforts coördinated, he feels that the individuals and the class gain in mastery of decision-making and in problem-solving. Learning in groups is said to satisfy individual needs so that energies may be devoted to solving achievement problems. Indeed, Thelen tends to emphasize not only the individual's gains in the mastering of content, in the accepting of self responsibility, in the motivating of learning, but also stresses the acquiring of group process goals in group feeling and in the understanding of democratic adjustments.

With the current interest in the group and in group dynamics, however, little has been said about what makes a group *a group* as distinguished either from an aggregation or, indeed, from an individual. The popular preference for the word "group" suggests a nuance of an organic unit or a cohesive whole. In the literature, however, the structure of the group is ill-defined and ambiguous. In the references that have been reviewed, the structure of the group, as used in elementary school teaching, may be broadly dichotomized, (1) as that in which the group is considered more in terms of method of instruction than in terms of its organization; and (2) as that in which the group is referred to in terms of rôle and interaction patterns.

The first obviously does not define a group at all, but gives the flavor of the classroom climate in relation to relative limits placed on child activities both by teacher demands and decisions as well as by the environmental constraints and limitations considered necessary for teaching and learning. Essentially, group structure merely —or really suggests the delicate balancing between the Charybdis of permissiveness and the Scylla of control. The second category asserts that there are differences between an aggregate and a group such as the fact that a group's members have common needs, interests and purposes. This, too, fails fully to specify a group, for in the schoolroom in so far as members have different needs, interests and purposes, they would form different groups. Perhaps, the closest indication of the differentia is in the process of interaction among the members of a group. Such an approach to the group leads to difficulties in the appraisal of and even the description of the communications and transactions among the members. Yet, the groupness of any group must represent some sharing of experiences, knowledges, feelings and purposes. Perhaps groupness, or as one writer terms it, "we-ness" is in the sharing of resources, be they intellectual, emotional or material, to accomplish a group purpose or several group purposes.

Even now, however, there is little or no evidence that suggests the degree to which groups achieve any one group purpose, or the extent to which the members as individuals gain from group experience in such objectives as social skills and knowledge, or, more significantly, in the ability to transfer such social gains to other groups. It is too early to expect that the effect of the group experience upon the school children can be appraised. At least, for research, there are hypotheses that should be tested such as: (a) any group benefits are shared equally by all individual members; (b) individual learning is maximized through the group; (c) group experiences strengthen the individual's need to learn; (d) group experience allows the individual to develop at his own rate to maturity; and so on.

It is my belief, however, that it will be very difficult to design research in the effect of the groupness of the group upon the behavior of its individual members largely because the group is so ill-defined.

In the course of summarizing the literature on experimentation with groups in problem solving, we became aware that the psychologist used the term "group" in at least five different meanings. This is not strange. For the lexicographer considers a group to be a "number of persons (or things) gathered together and, thereby, forming a recognizable unit; a cluster, an aggregation, or a band" or to be a "number of persons classified together because of common characteristics or community of interest." The difficulty lies in the fact that the dynamic character of the group is lost in the dictionary's definition. Psychologists tend to think of the group as made up of members, each of whom is in close personal relation with every other member. The group implies interaction and communication among two or more individuals whether with, or without, a designated leader but with a tradition of having learned in time and with mutual experiencing to work together. Obviously, such a coöperative amalgamation of individuals who have progressed upward through the stages of organizing, of assaying each other's resources and personalities and of establishing channels of communication into the more significant and higher stages of mutual reinforcement of one another's contributions to solving a problem is not easy to find. Except as described in terms of process observation, the traditioned group has not yet been studied. It is reasonable to assert that such groups have not been utilized in research in groups, either at the elementary or high school levels. And, with the exception of Darley's work with sororities, such groups have not been the basis of research with college students.

The psychologist, however, tends to generalize anticipated values of traditional groups to very many enterprises. Indeed, the psychologist and the educator have failed to recognize that in their research activities, they must, necessarily work with groups that are not traditional groups. All of you are familiar with the so-called *ad hoc* group. The *ad hoc* group is usually established by some outside agency (viz., the college teacher as experimenter) to work with, or solve some externally specified task, under the instruction that they do so coöperatively. Such an *ad hoc* group is at least one step away from the more dynamic group. The *ad hoc* group is an artificial assemblage made for the duration of an experiment. Such a group rarely is expected to, and rarely does, continue to develop groupness.

Under an Air Force contract, our research staff has attempted to test the hypothesis that problem solving by *ad hoc* groups or staffs is superior to problem solving of the random individual. Two widely different kinds of problems have been used with two different populations. One type of problem, designated as the field type, is like the familiar O.S.S. task of requiring a cadre of men to cross a mined road without attracting the attention of the enemy and without traces of their route. A series of such problems was given to *ad hoc* teams of college undergraduates to do. In general, the solutions by teams were markedly superior to those of random individuals. Such a result would have been anticipated from the published results of Shaw's well-known experiment in problem solving.

The other type of problem required the solution of a complex human relations difficulty such as taking steps to improve the morale at an Arctic Weather Station under the constraint that the mission of weather reporting must be accomplished. This type of problem was used with officers in the Air Force in the ranks of captain through colonel. The *individual* solutions tended to be as good or better than the average of *ad hoc* staff solutions.

Is it possible that college students work more coöperatively than officers; or is it likely that staffs or teams are superior in those problems for which there is a knowable solution; or is it possible that the responsibility of command makes each member equivalent to a staff? The answers are, of course, unknown. Certainly children in schools today are expected to participate in, and with, groups. Perhaps the future holds bright promise for group decisions and group problem solving if such experience develops a common element of groupness which is transferred positively to all forms of later group experience.

Edward L. Thorndike would certainly have wished to evaluate the attainment of the objectives. At Teachers College, three attacks are being made on the nature of the differences in the solutions to problems by groups and by individuals. The first is to compare *ad hoc* groups with random individuals; the second to compare *traditional* groups with *ad hoc* groups; and the third is to create mathematical models to account for the group superiority. As a matter of fact, Dr. Solomon and I have developed two such models to consider group behavior in the solution of the Eureka or knowable solution problem. The models are expressed in terms of two hypotheses: Model A, that group superiority is a function only of the ability of one or more of its members to solve the problem without taking account of the interpersonal rejection and acceptance of suggestions among its members; Model B hypothesizes that group superiority is a function only of the pooled abilities of its members. Model B implies that each problem is solved in two or more stages.

The use of Models A and B on the Marjorie Shaw data as well as the data for the O.S.S. field type problem suggests the *acceptability* (but not the proof) of the formulation that the group is superior to the individual *only* because any group has a higher probability of having among its members at least one solver, or several solvers who can independently solve the various stages of the problem. If this be verified by further researches with a broad range of groups and problems, it reopens the question of the "groupness of the group" and the enduring result as a group activity over fairly long spans of time.

Our own experimental evidence contrasting problem solutions by *ad hoc* teams made up of individuals, who had solved the *same* problem previously, supports the generalization that group superiority may be a reflection of the ability of the best one or two members of the *ad hoc* group. In another experiment, an attempt was made to contrast the problem solutions made by "traditioned" and by "*ad hoc*" teams. The traditioned teams comprised members who worked together at the Air University on all group activities as a staff for three months. The results lead to the inference that the "three-month traditioned group" is equivalent, but not better than, the *ad hoc* group. It must be emphasized that experimental, "traditioned" groups originally had been designated by the authoritarian procedure of selecting individuals from the class at random to make up the group. Hence the "traditioned" group may not have had in it members who would have developed the group cohesiveness that is so essential to a dynamic group with groupness.

No interest group, as sophisticated as educational psychologists, would be tempted to overgeneralize these results. First, the subjects in our experiments were adults who may have overlearned those ways of working out problems that stressed the individual; second, the solutions did not have to be carried into action. While the individual adult, student or officer, may have made a decision or solved a problem as well as or better than the group, this does not suggest that the individual solution must, necessarily, be carried out as well as a group solution or decision; and third, no attempt was made to ascertain what each individual as individual got from the group process of deliberation and closure.

The three limitations to generalizations about individual over group superiority should suggest that educational psychologists must develop hypotheses and the means for evaluating them; the attainment of curricular objectives not only for relatively immediate consequents but also for long time transfer to the tasks and problems that living affords. Indeed, the groupness of the educational psychologist lies not so much in the affiliations with other interest groups; but in the common purposes and needs to get at the facts, generalizations and applications that will lead to better teaching and better learning among themselves and for those that must learn. Groupness for educational psychology is a function of our communications and transactions.

A RELATIVISTIC DEFINITION OF STIMULUS-RESPONSE

MORRIS L. BIGGE

Fresno State College

Stimulus and response are terms which have become deeply embedded in the vocabulary and the ideas of educational psychology. In recent years the ideas represented by these words often have been made to correspond to the concepts of cause and effect as used in the physical sciences. Psychologists, in their attempt to avoid assumption of any metaphysical purposiveness, have posited a mechanistic behavior devoid of purpose in any sense, teleological or otherwise.

Current mechanistic educational psychology shows the influence of attempts to derive directed activities without assuming directed dynamic factors in human behavior. Behaviorists and connectionists have attempted by means of concepts, such as "goal gradient" or "law of effect," to recognize directed factors in behavior without surrendering a basically mechanistic position. Inasmuch as it is coming to be generally recognized, in educational psychology, that an atomistic stimulus-response idea is not adequate for describing human behavior, the tendency of many contemporary psychologists who continue to hold to a mechanistic outlook is to speak of "molar behavior" and of "total responses to patterns of stimulation." Educational psychologists may recognize behavior as molar, speak of situations and patterns of response, and emphasize "belongingness"; yet, as long as they maintain their allegiance to some modern version of associationism such as behaviorism or connectionism, they are clinging to a frame of reference fundamentally different from, and incompatible with, that of a truly relativistic field-theoretical approach.

Gates states, "The term *connection*, bond or tendency, does not refer to any neurological concept or agent at all. It refers to a functional relation between a situation and a response. It refers to an observed phenomenon and implies no neurological correlate. It implies nothing except that there is an observed tendency for a situation to be followed by a response." (*6*) He later continues, "Stimulus-response is a psychological, not a neurological theory." To one relativistically oriented, "psychological" means in harmony with the logic of a growing mind. Thus in the interest of consistency, there would seem to be a need to redefine stimulus as something other than a "situation to be followed by a response." For a stimulus to be a stimulus a response must be going on.

A prevalent current practice among educators of defining learning as change in behavior reflects or implies mechanistic definitions of stimulus and response. Observable *results* or *outcomes* of learning are interpreted to be the actual learning. It is true that when a person has learned or is learning, his behavior probably changes. However, it does not follow that, for learning to take place, there must be an observable change in behavior. Change of behavior is an indication that learning has been or is taking place, but the change of behavior is not the learning. Learning is more adequately defined as development of insight or understanding.

Traditionally, within patterns of mechanistic psychologies, insight has been defined as a special kind of learning. Insight as used by a relativist does not necessarily imply, "some penetration into the (absolutely) true nature of things." (*9*) It means the way one senses or sees things in a situation. It is a basic sense of pattern or relationship (*1*). Insight, thus, might be defined as a "sense of belongingness"—should belongingness be construed in non-mechanistic fashion. However, it cannot be defined as a sense of belongingness that is gained through repetition, in that repetition as such is not essential to the learning process. A more adequate theory would define all learning, at least so far as human intelligence is involved, as development or modification of insights. Insights are instrumentally, but not absolutely, true. They serve as accurate or dependable guides for action.

A relativistic field theory differs from any mechanistic theory of human behavior in that it assumes that intelligent behavior is purposeful. This assumption, however, is accompanied by no attempt to assert or deny existence of teleological forces. It simply means that a person, behaving intelligently, acts as if he were trying to do something; and that, if one is to predict accurately what a person will do in a given situation, he must envision what that person is trying to do.

Two basic tenets of a field theory are that behavior is to be derived from a totality of contemporary co-existing factors, and that these are to be considered in a dynamic field of such nature that the state of any part of a field depends upon every other part of that field. A relativist, in the study of intelligent human behavior, assumes that each person is unique in purpose and in experiential background; and that the reality upon which one bases intelligent action consists of that which he *makes* of the objects around him. In any field—science, social relations, morality, and even religion—each individual is the maker, not of *the* world but of his *notion* of the world. For psychological theory to be most applicable to the problems of modern education, in light of the present stage of scientific development, mechanistic patterns of thinking must be superseded by conceptualizations subsumed under a relativistic outlook. This relativistic outlook in educational psychology would reflect the influence of field theories such as those developed by Kurt Lewin, Boyd H. Bode, and kindred academic spirits, and would harmonize with the philosophic outlook of John Dewey.

If the terms stimulus and response are to be retained and incorporated into a field psychology, developed within a relativistic frame of reference, their meanings must be carefully redefined to harmonize with this pattern of thinking. Field psychology entails a situational point of view as opposed to the historical approaches of most earlier psychologies. Behavior depends on neither the future nor the past but on the present field. To predict behavior of a person one must take into consideration his total contemporaneous life-space which includes the person, his environment, his insights, and his goals.

TREATMENT OF MOTIVATION

A relativist in educational psychology assumes that intelligent human action is purposive and must be interpreted in light of the goals it is designed to achieve. An individual acts in such a way as to achieve his goal or goals—satisfy his wants or desires—in the quickest and easiest way that he comprehends or senses as possible under existing conditions. When a person behaves intelligently, he acts as if he is pursuing a purpose and as if he has some foresight as to how the purpose is to be achieved. When one has a purpose, he displays a capacity to refer present conditions to future results, as well as future consequences to present conditions.

Man's ability in abstraction and conceptualization enables him to define his purposes symbolically, through use of articulate speech. When a person is motivated toward doing something, his description of the situation is that he wants or desires to do it. He consciously carries activity forward to a goal through a process within which he constantly searches out conditions for the next step, all along the way. Moreover, a relativist who takes his relativism seriously sees continuity in the nature of learning throughout the animal kingdom. He observes purposefulness of behavior in adults, in children, and in lower animals, each on his own intellectual level.

Conventional laws of psychology are based on statistical predictions. Statistical predictions may apply to the average of children or to the typical behavior of an age-group, but they do not predict what a certain given person will do in a specific situation. Objectivity in psychology demands representing the field adequately and accurately as it exists for the individual in question at a particular time. Thus, in a sense, to be objective in psychology one must be subjective. One must look at things as the person being studied looks at them. In Lewin's basic formula, $B = f(P,E)$, behavior is produced by the totality of co-existing facts. This totality is a psychological field. This psychological field is not defined in terms of mere consciousness or awareness. Rather, it is a dynamic whole of such nature that a change in any part affects other parts, and every change depends upon the whole. Behavior is a f—function of P—psychological person and E—psychological environment.

Mace (*8*) has developed some extensions of Cannon's theory of homeostasis; (1) what is maintained or restored is not so much an internal state of an organism as some relation of an organism to its environment, (2) discovery of new equalibria is a more apropos description than restoration of equilibrium, (3) in the end state of a goal-directed process something is discovered which has not been experienced before or new goals are created. In attention, tensions are set up when we are presented with something incomplete, unclear, or unfamiliar. A task is presented or a problem is posed. Processes ensue which terminate in completion of the task or solution of the problem. Something is discovered and in discovery tension is relieved.

A relativist, in defining psychological terms such as stimulus and response, assumes a bi-polar relationship of person and environment. Instead of thinking of an environment (stimuli) impinging upon a physical organism, he assumes that in any situation a person acts in accordance with the psychological field which he perceives. This field is not defined in "objective," physicalistic terms, but in the way it exists for a person at a given time. A thing is perceived only as a relationship among an object, a viewer, and the background of the situation. In a psychological field, rather than passive behavior as a mechanistic response to an activating environmental factor called a stimulus, there is postulated an actively interactive experience—conscious behavior. Experience consists of a person trying something and simultaneously undergoing or noting the consequences of his attempt. "To 'learn from experience' is to make a backward and forward connection between what we do to things and what we enjoy or suffer from things in consequences." (*4*)

INTERACTION—PASSIVE OR ACTIVE?

Interaction is a popular term in current treatises on educational theory. When one searches for an incisive definition of the concept he finds that, in describing person-environment relationships, the word is used in two quite different senses. It may be used to denote a passive process within which a person or an organism is stimulated or modified by the environment and, in turn, the environment is modified by the organism. Each, organism and environment, is assumed to be passive while it is being acted upon by the other; e.g., a man is bitten by a dog, then the dog is kicked by the man. A stimulus from an object impinges on the end organs of a passive organism. The organism then responds. The temporal sequence is stimulus-organism-response. Since it is recognized that a human organism can modify its environment—change something outside of itself—the process within which the environment influences is called interaction or the interactive process.

Interaction has been used in a second (relativistic) sense by John Dewey and others to describe a process in which a person actively appraises his environment as contrasted with his being passively stimulated by it. "An experience is always what it is because of a transaction taking place between an individual and what, at the time, constitutes his environment." (*5*) One's environment consists of those conditions which interact with personal needs, desires, purposes and capacities to create the present experience. Experience includes an active and a passive element peculiarly combined. "On the active hand, experience is *trying*. . . . On the passive, it is *undergoing*. When we experience something we act upon it, we do something with it; then we suffer or undergo the consequences. We do something to the thing then it does something to us in return: such is the peculiar combination. The connection of these two phases of experience measures the fruitfulness or value of the experience. Mere activity does not constitute experience" (*4*).

The concept of active interaction in experience involves a description of situations. For a person to live in a world he must live in a series of situations. One experiences through acting and perceiving the consequences of his acts. Every experience both takes up something from those experiences which have gone before and modifies in some way the quality of the experiences which follow. It influences in some degree the objective conditions under which future experiences are had. When a child learns a thing, he determines to some extent the environment in which he will act in the future.

In Lewin's basic formula $B = f(P,E)$, the state of the person and that of the environment are not mutually independent. In a sense, the state of the environment depends upon the person and, conversely, the state of the person depends upon his environment (*7*). It is impossible to derive the psychological processes in the life-space without including, in the representation, changes in both person and environment. A bi-polar relationship of person and environment determines behavior. The psychological environment is the environment as it exists for the person. It is not identical with a person's physical surroundings. Everything in which, toward which, or away from which a person can perform psychological locomotion is part of his psychological environment. "One of the basic characteristics of field theory in psychology, as I see it, is the demand that the field which influences an individual should be described not in objective physicalistic terms, but in the way in which it exists for that person at that time. . . . To describe a situation "objectively" in psychology actually means to describe the situation as a totality of those facts, and of only those facts, which make up the field of that individual" (*7*).

A relativist, taking available data into consideration and harmonizing it to the best of his ability, concludes that a concept of active interaction is highly fruitful in a scientific study of human behavior. He recognizes that there are incidents in which one is passively influenced by an object from outside his organism; e.g., when a man not cognizant of any danger is shot from the back. However, whenever there is intelligent behavior, active interaction is highly descriptive of the situation. One cannot derive behavior teleologically from the future, neither can one derive it from the past. A psychological event is always a result of the interaction of many factors and is to be explained in terms of apparent relations.

Psychological person and psychological environment function as sub-wholes of a psychological, contemporaneous field or life-space. If stimulus and response are to be retained as descriptive terms in this frame of reference, stimulus suggests environment-centeredness whereas response refers to the functioning of cognitive and manipulative abilities of a person.

STIMULUS AND RESPONSE DEFINED WITHIN RELATIVISM

A relativist does not insist on discarding the terms, stimulus and response; neither, however, does he use them in the mechanistic sense of their origin. Behavior, for him, is not a passive response to an activating environmental factor called a stimulus. He sees little basis for assuming that "a crucial interval of delay occurs" (*3*) between the psychological stimulus and the psychological response. He redefines stimulus and response to mark a distinction within a larger field. Since an entire field operates as a unit, stimulus and response cannot be separated from each other temporally; they can be contrasted only in terms of function. To repeat, for a stimulus to be a stimulus a response must be going on. One does not first see objects and then respond to them—he sees objects in terms of the response he makes to them. "The reason why we speak of a stimulus at all is that the coördination or situation is *inadequate*; there is a drive or pressure towards a better coördination or adaptation. The stimulus is that phase of the situation which requires to be made more definite or explicit; the response is constituted by the reactions which create the need for a more adequate determination of the conditions for further activity. A stimulus, then, is a stimulus, in a psychological sense, only as long as there is this need of greater definiteness. When this definiteness is achieved, there is no longer any purposiveness at this point; the resultant action becomes an element in the next coördination" (*2*). Events such as seeing and hearing take place because, already, action is going on. The response is not a consequence of a sensory experience but is an antecedent or condition of it. Thus the reflex-arc hypothe-

TABLE 1—KUHLMANN-FINCH AND DAVIS-EELLS MEAN IQ SCORES AND STANDARD DEVIATIONS FOR EIGHT GROUPS OF ELEMENTARY PUPILS FROM THREE LARGE CITIES

	Kuhlmann-Finch		Davis-Eells		N	Grade
	M	σ	M	σ		
Knoxville, First Grade						
High socio-economic	108.59	8.1	103.07	15.1	27	1
Low socio-economic	95.37	11.1	86.37	13.6	27	1
Nashville, Third Grade						
High socio-economic	104.17	11.2	104.47	11.8	30	3
Low socio-economic	91.35	12.2	90.62	12.4	26	3
Nashville, Fifth Grade						
High socio-economic	107.90	12.9	113.85	11.3	28	5
Low socio-economic	95.65	12.5	99.63	9.8	30	5
Memphis, Third Grade						
High socio-economic	108.35	14.4	108.64	16.6	109	3
Low socio-economic	100.05	12.5	96.62	13.7	117	3

the guidance directors[1] in Memphis and Nashville. Miss Norene Bogart administered the tests in Knoxville in conjunction with her master's thesis.

The hypothesis was tested that differences in mean scores on the two tests would not be statistically different for the low or high socio-economic groups. To substantiate the claims made for the Davis-Eells Games, the scores made on this test by children of low socio-economic status would have to be significantly higher than on other tests, whereas children from high socio-economic homes would probably fall below on the Davis-Eells as compared to other tests.

FINDINGS

Table 1 shows the Davis-Eells and Kuhlmann-Finch mean I.Q. scores and standard deviations for all of the groups tested. To test the hypothesis stated above, analysis of variance was used (*3*). The results of this analysis are presented in Table 2. Bartlett's test was used to demonstrate homogeneity of variance.

As may be seen from the F ratios given in Table 2, the mean

---

[1] The authors wish to express their appreciation for the coöperation of Dr. Ellen Piers of Nashville and Miss Catherine Moores of Memphis in the gathering of these data.

TABLE 2—ANALYSIS OF VARIANCE OF TEST SCORES FOR FOUR GROUPS OF HIGH AND LOW SOCIO-ECONOMIC SUBJECTS ON THE DAVIS-EELLS AND KUHLMANN-FINCH

Source of Variation	Sum of Squares	df	Mean Square	F
Between tests	379.71	1	379.71	6.352
Between socio-economic levels	25,668.22	1	25,668.22	429.38*
Between pairs	4,759.36	3	1,586.45	26.53†
Interaction: test and socio-economic level	300.15	1	300.15	5.026
Interaction: test and pairs	2,545.44	3	848.48	14.19*
Interaction: socio-economic and pairs	1,205.03	3	401.67	6.719
Triple interaction	99.40	3	33.13	0.554
Within groups	46,153.19	772	59.78	
Total	81,110.50	787		

* Significant at the 5% level of confidence.

† Significant at the 1% level of confidence.

IQ scores on the Davis-Eells for the low socio-economic groups are not significantly higher than for the Kuhlmann-Finch. For the high socio-economic groups, the mean IQ scores on the Kuhlmann-Finch are almost the same as for the Davis-Eells.

These data would seem to support the tenability of the hypothesis that no mean differences exist for low or high socio-economic groups on the Davis-Eells Games compared to the Kuhlmann-Finch Intelligence Test. The claims made for the Davis-Eells Games are not supported in this study, which leaves unsettled the question as to whether there are basically no mean differences in reasoning or intellectual ability for children of high and low socio-economic status if tests could be developed free from cultural learnings. The present writers know of no way to test this assumption, but they believe that the position that mean differences in reasoning ability exist between low and high socio-economic groups is at least as tenable a position *(8, 10)* to take as the one adopted by Davis, Eells, Haggard, and their colleagues.

The F ratio between socio-economic levels in Table 2 is 429.38, which is significant at the 5% level of confidence. Thus, the mean scores on both tests are significantly higher for the high socio-economic children over the low socio-economic children. This is not a novel finding as it concurs with previous studies with older group intelligence tests (*6*). Since the claim cannot be made that the two instruments are measures of innate capacity devoid of environmental influences, the present authors recognize that the data do not constitute conclusive evidence for the hypothesis they have advanced.

In Table 2, the F ratios for the between pairs comparison is significant at the 1 % level of confidence. The data available do not permit an analysis to determine whether this significant difference is attributable to the differences in schools from the three communities or to factors associated with grade level. Lennon (*9*) has demonstrated that achievement at one grade level in a community does not permit accurate prediction for another level.

If Table 1 is re-examined, it might be noted that there are variations in mean IQ's from grade level to grade level and from community to community for each of the tests, and also between the tests. Such data need to be brought to the attention of those test users who remain convinced that the IQ is a completely reliable concept irrespective of the test used, grade level tested, or norms obtained by the publisher.

Correlations between the Davis-Eells and the Kuhlmann-Finch were also computed in view of the following comments made in the manual for the Davis-Eells Test (*2*, p. 11).

"The correlations of Davis-Eells scores with scores on other group intelligence tests should not be expected to be as high as has sometimes been reported for group tests in the past. The present test is built around a specific definition of intelligence in terms of ability to solve problems of a kind important and interesting to children; most intelligence tests in the past have been constructed specifically to predict academic success in the ordinary school situation."

The median r for the ten groups was .63 which is not much lower than the coefficients of equivalence or stability reported for the Davis-Eells (*2*, p. 20–21) or the split-half reliabilities (corrected) for the Kuhlmann-Finch (*5*, p. 53). The unreliability of both instruments would probably account for most of the variance not accounted for in the correlations between the two tests. This would seem to be evidence contradicting the claims made for the Davis-Eells quoted above.

SUMMARY AND CONCLUSIONS

In an effort to test the claims made for the Davis-Eells Games with respect to the cultural fairness of the test, 194 children from high and 200 from low socio-economic homes in three cities were administered the Davis-Eells and the Kuhlmann-Finch tests. The data obtained did not support the claims for the Davis-Eells that its items are culturally more fair than items used in other group intelligence tests. The writers have questioned the appropriateness of the assumptions made by Eells, Davis, Haggard, Havighurst, et al., that mean differences in innate ability between children of high and low socio-economic status are insignificant.

REFERENCES

1) Allison Davis, *Social-Class Influences Upon Learning*. (The Inglis Lecture, Harvard University, 1948). Cambridge, Mass., Harvard University Press, 1948.

2) Allison Davis and Kenneth Eells, *Davis-Eells Test of General Intelligence or Problem-Solving Ability Manual*. New York, World Book Company, 1953.

3) A. L. Edwards, *Experimental Design in Psychological Research*. New York, Rinehart and Company, 1950.

4) Kenneth Eells, Allison Davis, *et al.*, *Intelligence and Cultural Differences: A Study of Cultural Learning and Problem Solving*. Chicago, University of Chicago Press, 1951.

5) Frank Finch, *Kuhlmann-Finch Intelligence Tests Manual*. Philadelphia, Educational Testing Bureau, 1953.

6) Ernest A. Haggard, "Some Effects of Practice, Motivation, Item-Content, and Form of Administration Upon Intelligenic-Test Performance: An Experiment." *Genetic Psychology Monographs*, 1954, 49: 141–186.

7) Ernest A. Haggard, *Techniques for the Development of Unbiased Tests in Proceedings 1952 Conference on Testing Problems*. Educational Testing Service, 1953. Pp. 76–82.

8) Irving Lorge, *Difference or Bias in Tests of Intelligence in Proceedings 1952 Conference on Testing Problems*. Educational Testing Service, 1953. Pp. 76–82.

9) Roger T. Lennon, Stability of Achievement Test Results. *Educ. and Psychol. Measmt.*, 1951, 11: 121–127.

10) Fred T. Tyler, Comments on the Correlational Analysis Reported in Intelligence and Cultural Differences. *J. Educ. Psychol.*, 1953, 44: 288–295.

THE DOMINANT ATTITUDES OF NEGRO TEACHERS TOWARD INTEGRATION IN EDUCATION

ROBERT T. AMOS

Howard University

The decision of the Supreme Court that segregation in public education is unconstitutional will create certain problems which will arise as two large groups of people are brought together on an equal status basis for the first time. These groups will bring with them into an integrated system a background of varied experiences. They will view each other largely in terms of their experiences in a segregated culture. As they move from a segregated system into an integrated one, established patterns of living will be disrupted and this disruption will provoke differing feelings and attitudes. These feelings and attitudes must be considered in programs designed to effectuate integration. This study was undertaken in order to discover the nature of some of these feelings and attitudes which are provoked in Negro teachers with reference to the Negro's general ability to compete and to adjust in an integrated system.

The basic approach is qualitative, that is, subjective responses were elicited from a number of students without structuring the situation in advance. The responses were categorized, thus permitting some quantification. The categories used were derived from a content analysis of the responses and were not imposed *a priori*. It was assumed that simple verbal behavior is indicative of attitudes and that a suitable picture depicting a Negro on the campus of a southern white college would elicit from the subjects certain attitudes which were essentially their own and which would not be tapped by use of objective techniques.

SUBJECTS AND PROCEDURE

Twenty-five Negro subjects participated in this study. They were enrolled in a graduate course in educational psychology and were pursuing the Master of Arts degree in Education. There were five males and twenty females whose teaching experience ranged from one to nineteen years. These individuals were graduated from sixteen different Negro colleges in seven southern states and the

District of Columbia. Inasmuch as this was an exploratory study, no effort was made to secure a representative sample of the universe from which the subjects came.

The cartoon technique was employed. The cartoon was chosen after several preliminary test situations proved it to be sufficiently unstructured to yield a variety of attitudes about integration in public education, and at the same time, sufficiently structured to elicit attitudes relevant to Negroes attending the Georgia Institute of Technology.

The picture, in black lines against a white background on eight and one-half by eleven paper, shows a male figure with a black face, white lips and a rather large mouth partly open standing in front of a building. He has on slacks, and the sweater which he is wearing has a "T" on the front of it. He has books under his left arm with the left hand in his trouser pocket. Under his right arm he carries a "T" square. A building stands in the background with a clock in the tower and TECH can be seen engraved on each side of it. Beneath the drawing, a musical note is in front of these words, "Ah's a ramblin wreck from Georgjah Tech."

The drawing was presented to each individual subject for examination during a regular class period. The drawing was then placed where it could be observed by each class member and the following statements were made:

"Let us assume the artist released this drawing after the Supreme Court ruled that segregation was unconstitutional in public education. Let us assume further that the student publication staff at Georgia Tech released it. Look at the drawing carefully and you will observe a Negro on the campus. Now write your reactions to this observation."

All of the students knew that the caption under the drawing was a dialect quotation from the Georgia Institute of Technology school song. No time limit was placed upon the writing, and after each student completed the task the essays were collected.

The essays were analyzed for statements which implied an attitude toward integrating the Negro in public education. It was found that the one hundred and twenty-four items fell into seven categories. The seven categories established to define the attitudes expressed by the subjects were: intellectual inadequacy, rejection, social inadequacy, stereotyped beliefs, acceptance, resentment and ambivalent attitudes.

To check on the keying procedure by the investigator, three judges trained in psychology and sociology were employed. Each judge was given the definitions of behavior and the items grouped under each category. They were asked to check on the keying procedure in terms of the definitions of behavior, the items assigned to each category according to the attitudes implied, and each item listed under the category which did not parallel the definition of behavior.

The judges accepted the definitions of behavior established and the general body of items listed under each category except that of resentment. They considered the items listed under resentment too similar to those under rejection to form a separate category. Therefore, the body of items listed under resentment was placed under the category of rejection. Twenty-four miscellaneous items were eliminated because they were judged by the jury not to parallel the definition of behavior under which they were listed. Thus, ninety-six items were retained in the final classification under six categories in the order of frequency as shown in Table I.

PRESENTATION OF DATA

To illustrate the attitudes categorized in Table I, the statements in each area will be described in more detail. Twenty-nine per cent of the items fell into the category of intellectual inadequacy. These items expressed the idea that the Negro was deficient in the special abilities which are necessary to become a successful engineer. Two levels of inadequacy were indicated. The first level expressed the idea that the Negro is attending Georgia Tech, but he is having difficulty in accomplishing the academic tasks. In fact, they doubt if

TABLE I—CATEGORIES OF ATTITUDES EXPRESSED TOWARD INTEGRATION BY RANK AND FREQUENCY OF ITEMS

Category	Frequency of Items	Per Cent
Intellectual inadequacy	28	29.0
Rejection	21	22.0
Social inadequacy	15	15.6
Stereotyped beliefs	13	13.5
Acceptance	11	11.5
Ambivalent attitudes	8	8.0
Total	96	99.6

he has the ability to perform the tasks required in the school of engineering on a par with the white students. His inability to accomplish these tasks is expressed in such statements as the following: "He cannot master the work to become an engineer at Georgia Tech." "He has the tools and books necessary to do school work, but he does not accomplish it." "The school is so rigid that the Negro will have difficulty in trying to be a successful student." "The Negro has to go from place to place with his books and equipment realizing that every minute is needed for study."

On the second level of intellectual inadequacy, the Negro has completed the academic requirements at Georgia Tech, but he has not acquired the behavior expected of an adult engineer in our culture. The items on this level depict him as not accomplishing much after graduation, as an unfinished product, as not being able to fit in any job situation very well, and as not making any contribution to the school. The attitudes expressed here are not surprising since engineering is a field in which Negroes have only recently been permitted to enter.

Twenty-two per cent of the items fell in the second category expressing rejection. These items are concerned with not being wanted or needed in the social situations considered to be normal school activities. It is felt that the Negro would be permitted to attend Georgia Tech, but he would be ostracized and humiliated in his school experiences. There is a widespread feeling that the whites resented the Negro's intruding in their schools and that he would be made to feel that he did not belong there. The attitudes expressed in this area are: "The Negro will attend Georgia Tech, but will be socially ostracized by the white group." "After the person tried to enter school, his experiences were not wholesome so he is leaving, confused and displeased." "Negroes are not wanted at Georgia Tech." "White people in Georgia are not willing to accept Negroes in white schools."

The third category, social inadequacy, includes fifteen and six-tenths of the items as shown in Table I. These items convey the idea that the Negro is unable to employ the necessary social skills for the maintenance and enhancement of self in an integrated system. Thus, as the Negro moves from the segregated schools into the integrated system, the lack of social skill in mixed racial groups will foster self-consciousness rather than a feeling of confidence. Throughout this category the attitude was implied that the Negro was unstable, lacked self-control and did not have the skills expected for social living to feel competent as he moves out of a segregated culture. These views are variously expressed: "He is unstable and does not have much control over himself." "It depicts him as one who has not been able to find himself and make the necessary adjustment." "They feel when he enters, he will not be able to cope with the situation." "It is hard for him to adjust at Georgia Tech."

"Stereotyped beliefs" was the fourth category; these beliefs included thirteen and five-tenths per cent of the items which referred to standard conceptions of the attributes portraying supposedly distinctive traits of the Negro. In this area the body of items included statements similar to these: "He will attend classes looking very shabby." "He is a typical Negro—wide nose and big lips." "He is unkempt and happy." "The Negro boy is gleefully singing." "Negroes are going to 'overflood' the school."

The fifth category includes items which express the attitude of acceptance. Eleven and five-tenths per cent of the total items are in this category. These items express an attitude which is the reversal of the patterns of behavior expected of a Negro in a segregated culture. The Negro is viewed as attending integrated schools on an equal-status basis with whites. Morever, those in this category felt that the removal of segregation in education will result in the disappearance of other restrictions which will permit the Negro to be an individual in integrated schools and permit him to participate in those activities which are important to him. In this category some of the items are: "He feels a new freedom." "He feels he can combine his social aspect of life with the academic aspect and be accepted in both." "He feels he can attend Georgia Tech on an equal status with whites." "His dress seems to indicate that he feels he can proceed now on his individual merit rather than because he is a 'proper' and 'correct' Negro."

The final category, ambivalent attitudes, which includes less than 10 per cent of the items, is characterized by the simultaneous existence of contradictory goals of equal importance to enhance self. The conflict is whether the Negro would rather remain in a pleasant situation where the educational facilities are meager for Negroes or go to integrated schools which are better equipped, but where his social relations would be uncertain. There appears to be a feeling that his status would be enhanced when he attends the integrated schools, but he is faced with the problems of uncertainty of the white group's reaction to him. These feelings were expressed in such items as: "He is deciding whether he should stay and learn what Georgia Tech has to offer or go to a more pleasant surrounding with fewer educational facilities." "The boy desires to take advantage of the engineering course, but he also desires to be away from the unpleasant environment." "He thinks he would feel better if he had remained in a segregated school." And, "The artist is portraying the boy's reluctance to go to Georgia Tech even though he has permission to go."

SUMMARY

When the subjects in this investigation interpreted the situation depicting the elimination of segregation in public education, six kinds of attitudes were categorized. The complete array of items listed under each category was not given in the body of this paper but samples of these items were listed as well as the percentage distribution and the number in each category. It appears that the samples of these items and the percentage distribution will suffice to suggest the kinds of attitudes evoked when Negro subjects reared in a segregated culture perceive this change. However, when the items listed under the categories of intellectual inadequacy, social inadequacy, rejection and stereotyped beliefs are examined, it is interesting to observe the negative aspect of the attitudes. It would appear that an individual who holds such negative attitudes as these would enter an integrated school system with inhibitions which would prevent him from competing on an equal status basis with his peers. Moreover, the attitudes expressed in the ambivalent category seem to have this same negative element. In this category the high esteem which the Negro holds for the majority group and the desire to enhance his status by attending integrated schools can be detected, but he also expresses the feeling that he is most comfortable with the minority group of whom he is most critical.

On the other hand, attitudes were expressed to indicate the subjects were adequate to accept the elimination of segregation in public education. It appears that the individual's concept of self and the expectations demanded of him by the elimination of segregation were in harmony. In the new situation the Negro felt he would be accepted on an equal-status basis and would be competent to participate in those activities which were important to him in an integrated system.

If the roots of democracy are so multitudinous and if their beginning is so early in life, then the selection of experiences that foster democracy require knowledge of the individual, knowledge of the effects of experience, and applications of these knowledges that should be conspicuous for their wisdom. Unless educators are to abandon their responsibilities they must search for such roots of democracy. Otherwise their selections among learning activities will not be superior to those that might have been made by lay intuitions.

The matter reported here adds something to our meager knowledge of democratic attitudes as they are related to the experiences of college students. The data here are part of a larger project directed toward the measurement of attitude. *A Coöperative Study of Evaluation in General Education* undertook as one aspect of its work to measure a democratic attitude of respect for the worth and dignity of every individual. In the development of the instruments there were investigations of many kinds. Two of the supporting investigations at Western Washington College of Education were directed toward description of democratic students in terms of competencies, life experiences and other identifiable characteristics.

A preliminary study sought to identify a sample group of students who displayed democratic attitudes and a sample group with non-democratic attitudes. Thirty faculty members, in ten committees of three members each, selected students who exemplified extremes in the presence or the absence of a democratic attitude of respect for others. The committees were forced to derive their own definitions in order to avoid the structuring which might result from the biases of any explanation. They were also asked to record data substantiating their judgments. The list submitted by these committees consisted of 67 males and 42 females. Of these 109 cases, seven students were listed in both groups and were therefore dropped from the sample. Students who were listed as possessors of the democratic attitude were found to have significantly distinguishing characteristics which suggest that differences exist other than those that are dependent upon attitude alone.

A second group of investigations used an advanced trial form of the *Problems in Human Relations Test* (3) and dealt with data from the responses of 1068 unselected college students. The sample consisted of 373 freshmen, 480 sophomores and juniors, 215 seniors;

612 men, 456 women. The *Problems in Human Relations Test* (PHR) is a partially disguised, multiple choice attitude test. It was scored on a major or "Democratic" D scale and on four sub-dimensions; "Hard-boiled Autocrat" (HA), "Benevolent Autocrat" (BA), "Laissez-faire" (LF), and "Resort to Expert" (RE). The sub-scales were designed to describe the nature of low "Democratic" scores. All of these scores were compared with more than fifty variables including: sex, age, interests, intelligence, English competence, reading skill, father's occupation and education, type of home community, critical thinking ability, high school grade point average, and philosophic point of view.

The consistent relationship of certain variables with scores on the democratic scale warrants generalization. Other marginal and not always consistent relationships suggest explorations but are perhaps not essential to discussion here.

The most pronounced and pervasive attitudinal difference that affected all other variables was a difference based on sex. Faculty judgment and test performance demonstrated that much more democracy existed among women than among men. From a student body with a 3 to 2 male majority, faculty committees selected almost identical numbers of men and women as examples of democratic students. These committees named students who embodied democratic attitudes and also named students characterized by lack of this attitude. Of all the women named, 71% were in the positive or possessor group, whereas only 46% of the men were so characterized.

Data from the subsequent test records of more than 1000 students evidence strong female acceptance of democratic responses. The mean democratic score for women was 17.22; for men, 15.15 and the critical ratio of difference was 7.67, well above the .01 level of significance. Subsequently, separate norms were used.

The democratic limitations of male students are further defined in their autocratic and "laissez-faire" scores. A mean score of 4.76 on autocracy was significantly above the women's 4.00, with a critical difference of 6.49. Likewise, their "laissez-faire" mean of 1.60 was enough above the women's 1.45 to compute a critical ratio of difference of 4.84, demonstrating significance above the .01 level. It seems that masculinity is autocratic or at least non-participating but is not essentially democratic as defined by faculty committees or by the *Problems in Human Relations Test.*

Marked differences between democratic and undemocratic students were also found in academic competency. Students named in the positive group by the faculty were significantly superior in college grade points, in grade points for college science courses, in language skill as measured by the *Coöperative English Test* (Mechanics), and in academic capacity as measured by the *American Council Psychological Examination*. The significance of these differences is reported in Table I.

Later evidence based on a division of students by test scores rules out the possibility that the competency differences were really the basis on which democracy was rated. Correlations between democratic and competency scores yielded coefficients of .22 for reading, .38 for English total and .24 for the L score (ACE Psychological Examination). More striking relationships, however, were discovered when the lower quartiles of academic cömpetence were considered as predictors of limited democratic attitude. Limitations in English grammar or mechanics were related to high scores on the anti-democratic subdimensions of the *Problems in Human Relations Test*. The students in the lowest quartile of English grammar had a mean score of 4.55 compared with the upper quartile mean of 3.54 on the (HA) "hard-boiled autocrat" category, with a computed critical ratio of 3.43. In the "benevolent autocrat" category this comparison was 5.51 to 3.64 with a critical ratio of 4.96. Total English scores were also significantly related. When a division was made on this basis, students with less adequate English scored 5.68 on "hard-boiled autocracy," whereas the abler students scored 4.50 with a critical ratio of 7.27. The comparison on "benevolent autocracy" is between scores of 6.00 and 3.66 with a difference of 2.36, significant at the 5 per cent level.

The "benevolent autocracy" subdimension was later dropped from the test due to internal shifts. These data did, however, dem-

TABLE I.—COMPARISON OF ACADEMIC SKILLS FOR STUDENTS RATED ON DEMOCRATIC ATTITUDE BY FACULTY

	Democratic	Non-Democratic	C/R
Academic GPA	2.76	2.50	4.23
Science GPA	2.55	2.25	4.98
Gross ACE.	116.78	111.94	2.48
Coöperative English	66.94	55.70	2.99

onstrate that low English scores are related to the undemocratic responses on the attitude test. The consistency te the conclusion was further demonstrated by the positive relationship between total English and democratic scores. Students in the top quarter of English performance scored a mean 17.86 on the "D" (democracy) scale whereas the lower quarter scored 12.70. This ratio of 6.78 reaches significance beyond the .01 level.

Among men the top quartile of hard-boiled autocrats scored a mean of 41.17 on the Linguistic or L scale of the *ACE Psychological Examination*. The least hardboiled of the male student group had a mean L (linguistic) score of 61.57. The significant ratio of 6.86 was computed.

It follows that students classified as democratic by this test have somewhat superior English skills but even more characteristically students with low democratic and high autocratic scores have limited English skills.

A third constellation of evidences centers on age and college tenure. The results summarized in Table II demonstrate an increase in democracy with a decrease in autocracy as students move from the freshman to the senior year. A shift away from benevolent autocracy was also found and this change, like the others, was much more pronounced for men than for women.

A reduction of autocratic scores occurred in male students. The average age of students in the upper quartile of autocratic scores was 21.10 years whereas it was 24.47 for those in the lower quartile. This difference is significant at the five per cent level. The greater change in men may reflect the lower initial scores or may result from a later male maturation. No evidence on this issue was available.

TABLE II—MEAN ATTITUDE TEST SCORES OF STUDENTS AT VARIOUS GRADE LEVELS

	Males				Females			
	"D"	C/R	"HA"	C/R	"D"	C/R	"HA"	C/R
Freshmen	13.55		5.31		16.90		4.47	
		6.18		3.13		1.10		1.45
Soph. & Jr.	15.28		4.68		18.31		3.67	
		5.36		1.75		1.78		0.36
Seniors	16.78		4.30		18.66		3.58	

TABLE III—CHANGES IN MEAN ATTITUDE TEST SCORES, ACADEMIC YEAR 1951–52

Scale (Dimension)	Fall	Spring Re-test	r	C/R
D	16.49	17.33	.70	3.23
HA	4.66	4.01	.49	4.64
LF	1.48	1.30	.40	2.25

More direct evidence of increase in democratic attitude was found in a comparison of fall-quarter mean scores and spring-quarter mean retest scores for 227 students as reported in Table III. The change between test and retest clearly confirms findings that during college the democratic responses are strengthened and the autocratic and laissez-faire responses are reduced.

A fourth group of relationships centers around personality, interest and critical thinking. *The Problems in Human Relations Test* was keyed by several experts operating under different assumptions. The result was twenty-eight keys. Some were developed in accord with the twelve categories of Rosenzweig picture frustration responses (*4*). Others were based on value categories, a hostility continuum and other personality categorizations. Use of the keys contributed to the understanding of sex differences. Women were less willing to blame others, more willing to accept blame, and less willing to affiliate with the punishing person or group. Women tended to place group values high and to hold themselves responsible for the group. They were, furthermore, inclined to expect solutions from someone else and to resort to expert authority.

Beyond this sex difference there were significant relationships between attitude and some of the scores obtained through using the new keys. The significant items of correlation are reported in Table IV.

It appears that aggressive and tough-minded behaviors fit with a constellation of hard-boiled autocracy, laissez-faire and blame. Refusal to decide or attempts to escape issues were not characteristic of democracy but did preclude selection as undemocratic. It follows that subjects inclined to recommend referral, open or disguised, were not classified as undemocratic.

Male students with high linguistic scores on the Kuder Preference Record accepted experts and were benevolent at statistically significant levels. They also tended toward the inactivities of laissez-

TABLE IV—CORRELATIONS OF SIGNIFICANT "FACTORS" DERIVED FROM DATA COMPUTED USING 28 DIFFERENT KEYS

	H	L	E	e	I	D′
H	1.00	.54	.76*	−.35	−.04	−.80*
L		1.00	.52	−.70*	−.19	−.70*
E			1.00	−.42	−.13	−.71*
e				1.00	.10	.44
I					1.00	.08
D′						1.00

KEY: H—Hard-boiled Autocrat; L—Laissez-faire; E—Blame, hostility against a person or thing; e—Solution expected of someone else; I—Blame directed upon self—Self-punishing; D′—Disguised referral (approval seeking).

* Significance greater than the .01 level.

faire reaction. Women with high scientific interest scores rejected laissez-faire solutions. No other categories of interest score maintained consistent relationships with attitude scores.

A test of critical thinking (also developed in the ACE Coöperative Study of Evaluation in General Education) was significantly and positively related to democratic scores for both male and female subjects. Women in the top quartile on the democratic scale also did well on a test of critical thinking in the social sciences (developed in the ACE Study). Only the significant pieces of data are reported in Table V.

Several other data might be considered as a group. These are the biographical facts that describe the community from which the student came, and his father's occupation, income, and education. "Democratic" scores were significantly high for students brought up in villages. The mean score for them was 21.36. Town (2500 to

TABLE V—SIGNIFICANT ITEMS OF RELATIONSHIP BETWEEN ATTITUDE AND CRITICAL THINKING

	Male		Female			
	High D	Low D	High D	Low D	High B.A.	Low B.A.
Critical thinking	34.50	24.15	34.16	27.55	—	—
Critical thinking in Social Science	—	—	30.88	24.11	22.70	33.00

25,000), city (25,000 to 100,000) and big city (over 100,000) scores were much alike with means of 17.36, 16.73, and 17.14. Farmers' children had the lowest "democratic" scores with a mean of 15.82. The size of home community was related to "autocratic" rather than to "laissez-faire" or "resort to expert" scores on the attitude test. Farm, town, city and big-city students were all significantly more benevolent than those from villages. Of these, the farm students gave most autocratic responses. These data should not be interpreted as descriptive of the various populations because there is no assurance that equal portions of each group go to the college involved. The generalizations here are only for the school population and are not applicable in general.

The father's education was significantly related to men's attitudes. Men with democratic scores in the top quartile had fathers with an average education of 9.04 years, whereas those with limited democratic attitudes had fathers with 10.91 years average education. The reverse of this on the "benevolent autocrat" score with means of 10.91 and 7.90 demonstrates that men students were more benevolent as their fathers had more formal education.

Father's occupation and income were examined as possible predictors of democratic attitude. Women with high democratic scores had fathers who were statistically higher on census occupational classification than the fathers of other women. On the Resort to Expert scale high scores for women were significantly related to lower occupational level and high income. It appears then that, as is the case with men, daughters of successful parents may depend on experts, but that they are more democratic when the success of their parents is measured by occupational scales rather than financial scales.

There is some evidence that other variables are related to faculty choice of democratic subjects. An independent rater found that men, attractive in physical appearance, tended to assignment as democratic, whereas good-looking women tended to be labeled undemocratic. There are no evidences to explain whether these data reflect bias or whether unprepossessing men respond aggressively and attractive women learn early that they can afford themselves the luxury of undemocratic behavior.

There is limited indication that subjects selected as democratic tend to choose the middle course and perhaps behave like the average. The concentration in the sample of democratic students were affiliates of the central Protestant churches. The sample was too limited to be conclusive in interpreting possibilities of choice from smaller religious groups. The personnel records were reviewed and summarized by another rater. It appeared that vigor and aggression are assumed to be non-democratic though males rated as independent were associated with respect for others as shown by high "democratic" scores. These data appear to concur with those from the twenty-eight keys previously reported. Democratic subjects are perhaps a mid-group in behavior: they affiliate with majorities in church affiliation; they are not burdensomely dependent nor are they outspokenly independent of their advisors. It would be unwarranted to interpret these findings as descriptive of subjects. They must, however, be considerations in any description of characteristics judged to be democratic.

This verbal middle-of-the-road behavior has often been considered characteristic of democracy in any situation. Research on the effects of democratic practice in the home has resulted in statements such as the following:

"Democracy is characterized by a high level of verbal contact between parent and child, appearing as consultation about policy decisions, as explanation of reasons for the family rules, and as verbal explanation in response to the child's curiosity. Accompanying this flow of verbal communication is a lack of arbitrariness about decisions and a general permissiveness plus restraint on emotionality. . . .

Socialization by definition demands the development of contradictory aspects of the personality. Conformity to cultural demands is not easily obtained without robbing the child of that personal integrity which gives him a mind of his own and which supports him in his attempts to satisfy his curiosity and to carry out his ideas and phantasies in his dealing with the real world. . . ." (*5*)

CONCLUSION

A vital question in education, at any level, is, "Who shall be selected." Sometimes disguised but always present is the function of rewarding, encouraging, promoting, and advancing some persons over others. The criteria of choice might be expected to reflect the values of a society or some perversion of these values. In some situations, the criteria may be dictated by values that are characteristic of some approved, powerful or subversive group in the society. On these criteria, some will be advised to remain in school and others will not; some will be guided into professions and others will not; some will be given degrees and others will not; some will be recommended to employers and others will not.

The significance of this selection can scarcely be over-estimated. The cultural, economic and human directions are altered by the wisdom employed in deciding what qualities shall be approved and by the accuracy employed in detecting which of the people best exemplify such qualities. The studies reported here provide some insight into the selections as they operate in an American college.

The student selected as democratic by test score is also selected by faculty committee. The agreement is well beyond the .01 level of significance. The selectee is likely to be able, verbal, older than average, the son of a villager, a successful student who performs well on a test of critical thinking. His home is likely to be non-affluent with his father's income and education slightly below average. His affiliations and behaviors tend to be the popular and widely approved ones. He is probably less democratic than his sister.

It is possible that faculty select students on the basis of appearance, sex, academic performance or verbal ability but the fundamental fact remains that having been selected (whatever the basis) the members of this group do perform in a significantly different way on an objective test that has been designed to elicit democratic behavior.

The evidence supports previous findings in demonstrating the pervasive and inclusive facets of relationship between democracy and ability, democracy and skill, democracy and maturity, democracy and family characteristics, democracy and biographic facts and democracy and test responses.

The in-person consistency found in these data is important in social theory. The most democratic subjects were able, skilled in critical thinking, socially acceptable, academically successful and could probably be identified by a lengthy list of approved characteristics. These evidences may indicate a composite superiority of which democracy is a part. Some democratic subjects are characterized by high rating in all of these. Persons who had none of these excellences were not selected as democratic.

These data are productive of one other basic and exploratory consideration, "What kind of democratic person do we desire?"

Perhaps the subjects selected are normative rather than democratic. Perhaps they are selected because the testmakers and the committees share the popular supposition that goodness lies in the middle.

It may be that democratic college students are verbal, are restrained and do restrict their individuality; it may be that the test items are incapable of eliciting responses characteristic of a firmer and more vigorous democrat, or perhaps these selectees with their majority of women, with their techniques of referral, with their high level of verbal and academic competence, and with their permissive mildness, are the same persons who can be counted on if grim necessity requires them to take a firm stand for democracy.

REFERENCES

1) T. W. Adorno, *et al.*, *The Authoritarian Personality*. New York, Harper and Brothers, 1950, p. 971.

2) Bruno Bettelheim and Morris Janowitz, *Dynamics of Prejudice*. New York, Harper and Brothers, 1950, p. 178.

3) *Problems in Human Relations Test*. American Council on Education, 1951.

4) S. Rosenzweig, *Rosenzweig Picture Frustration Study*. Pittsburgh, Western State Psychiatric Hospital, 1944.

5) Alfred L. Baldwin, "The Effects of Democratic Practice in the Home." *Psychological Studies of Human Development* (edited by Kuhlen and Thompson). New York, Appleton-Century-Crofts, Inc., 1952, pp. 138–139.

THE CHARACTERISTICS OF POOR AND GOOD READERS AS DISCLOSED BY THE WECHSLER INTELLIGENCE SCALE FOR CHILDREN

HAROLD F. BURKS

San Gabriel School District

and

PAUL BRUCE

University of Iowa

INTRODUCTION

Of continued interest to educators is the problem of the poor reader. Every teacher has a number of pupils who do not achieve in this subject to the extent that their measured general mental ability would indicate. Essentially the problem raised in this study is as follows: Do most of these youngsters actually lack certain abilities or capacities basic to the reading process.

In the present paper, the hypothesis is advanced that the poor reader may be relatively weak in those parts of intelligence tests which resemble vital characteristics inherent in written language. If a pattern of abilities for poor readers can be found, the resulting profile should have implications for curriculum modification.

It was decided that the Wechsler Intelligence Scale for Children (WISC) (*6*) would be a profitable instrument for exploration of this area since it is administered individually and because the writers, in using this instrument, noted that poor readers had more difficulty with certain subtests.

Various research studies have attempted to show a pattern of mental abilities which characterizes poor reading skill. In an extensive study of 30 poor readers by Robinson, it was found that the Knox Cube Test[1] yielded the lowest mental age equivalent of the subtests on the Arthur Point Scale (*5*, p. 154). Graham (*2*) reports that for 31 poor readers who were tested on the WISC, statistically significant differences were found for these subtests:

---

[1] The Knox Cube Test is a non-verbal test of immediate memory span in which the task is to repeat a series of taps on four cubes.

Arithmetic, Vocabulary, and Coding were significantly low; Similarities was significantly high.

PROCEDURE

The subjects of this investigation included 11 good readers (one or more years above grade level on the reading section of the Wide-Range Achievement Test) and 31 poor readers (one or more years below grade level on the reading section of the Wide-Range Achievement Test), all of whom were given the WISC.

They ranged in grade levels from third through eighth grades. The 11 good readers were made up of 6 girls and 5 boys; the group of 31 poor readers contained 5 girls and 26 boys. All of the children had IQ scores above 90, therefore they reasonably could be expected to read at least at grade level. The average Full Scale IQ score of the good group was 117, and the average Full Scale IQ score of the poor group was 101.

RESULTS

In Tables I and II are presented the rankings of the WISC subtests according to the number of months deviation from the average WISC mental age equivalent score in the direction of

TABLE I—RANKINGS OF WISC SUBTESTS FOR 31 CHILDREN IN POOR READING GROUP BY MONTHS DEVIATION FROM AVERAGE TOTAL SCORE

* 1. P. Arrangement	+17 m.
* 2. Block Designs	+11
3. P. Completion	+9
* 4. Comprehension	+8
5. Object Assembly	+4
6. Vocabulary	0
7. Similarities	0
* 8. Information	−8
† 9. Arithmetic	−8
*10. Coding	−12
11. Digit Span	−13

TABLE II—RANKINGS OF WISC SUBTESTS FOR 11 CHILDREN IN GOOD READING GROUP BY MONTHS DEVIATION FROM AVERAGE TOTAL SCORE

† 1. Similarities	+20 m.
2. P. Completion	+12
2. Information	+12
4. Object Assembly	+5
5. Vocabulary	+5
6. Block Designs	+3
7. Digit Span	+3
8. Arithmetic	−2
9. Comprehension	−2
10. Coding	−3
11. P. Arrangement	−4

* These deviations were significant at the 1% level of confidence.

† These deviations were significant at the 5% level of confidence.

Levels of confidence were derived by a t-test for correlated data. (*3*, p. 137) advanced to retarded. This method of presentation was chosen for two reasons: (1) The extent of the deviation from the average total test score can be easily seen; and (2) the difference in average mental ability between the group of good readers and the group of poor readers is held constant so that comparison can be made in like terms on the basis of deviations from their respective average mental ages.

DISCUSSION OF RESULTS

The present writers believe that their results and those found by the researchers mentioned above indicate a pattern common to poor readers which can be understood in terms of Goldstein and Scheerer's concepts of concrete and abstract (*1*).

According to Goldstein and Scheerer, the concrete attitude is the surrender to experiences of an unreflective character.

> We are confined to the immediate apprehension of the given thing or situation in its particular uniqueness. . . . Thinking and acting are directed by the immediate claims which one particular aspect of the object or of the outer world situation makes (*1*, p. 2–3).

In other words, a person is dealing with his world in a concrete manner when he reacts to stimuli in his environment directly without reflection, conceptualizing, or symbolizing. For example, when driving, most of us learn to react to a red traffic signal by stopping. We might react to the color red when it is presented in other situations by always coming to a halt. We would do this if we were unable to be anything but aware of "redness" as being a peculiar property associated with danger.

In the abstract attitude, according to the same authors,

> We transcend the immediately given situation, the specific aspect or sense impression; we abstract common from particular properties; we are oriented in our actions by a rather conceptual viewpoint, be it a category, a class, or a general meaning under which the particular object before us falls. (*1*, p. 3–4)

In other words, a person reacts in an abstract manner when he mentally leaves the immediate stimuli and, with the use of symbols as a tool, forms concepts and generalizations about the experience. For example, while driving we see an alternate route to the same destination and have to decide which is the best way by means of considering various factors (distance, traffic, road construction, etc.); in so doing, we read symbols, call upon past experiences and knowledge, form conclusions—thus react in an abstract manner.

Applying these concepts of concrete and abstract to the results of the present study, it is essential to discuss the nature of the subtests which ranked significantly above and below the mean total test score.

In three subtests—Comprehension, Block Designs, and Picture Arrangement—which in the present study were found to be significantly high for the 31 poor readers, the question arises as to whether there are any essential factors in common. It is proposed by the present writers, that the essential factors in common to these three subtests include (1) relative lack of need for long or short term, symbolic memories; and, conversely, (2) the immediate availability of a structured stimulus.

While it has been considered by Mayman *et al.* (*4*) that the Comprehension subtest measures judgment and the Picture Arrangement subtest measures anticipation and visual organization, yet both of these employ unique situational stimuli which are immediately available to the subject at all times. They require associational cues from past experience but of a very commonplace nature. For example, in the Comprehension subtest, the item, "What is the thing to do when you cut your finger?", binds the subject to a common experience which involves relatively nonsymbolic associations. In a like manner, while the Block Designs subtest is considered to involve a visual analysis and synthesis of visual patterns, yet it can be seen again that the stimulus (i.e., the design cards) though not commonplace, is always at hand and therefore relatively few demands are made upon memory functions. It seems to the present writers that these points essentially place these three subtests within categories calling for less abstract behavior (as defined by Goldstein and Scheerer) than is needed for the other tests in the scale.

On the other hand, in Coding, Information, and Arithmetic, the three subtests which measured significantly low for the poor readers, it can be seen that (1) memory functions are most necessary; and (2) a given stimulus does not remain immediately available. It is fairly obvious that the Information subtest demands recall of specifically learned material; the subject either knows the answer or does not. The low score on this test might possibly be influenced by the fact that, being a poor reader, the subject has had less opportunity to acquire knowledge. The Arithmetic subtest, besides requiring the immediate memory of the presented problem, involves a long-term memory of the processes of arithmetical computations and then the actual carrying out of these processes mentally to arrive at a correct solution. The Coding subtest—the most outstandingly difficult subtest for poor readers—which, on the surface, may seem not to require memory, actually depends upon the memory function for efficient performance in this task. The subject, to do well, must become familiar with the nine symbols and associate a digit with its more or less meaningless symbol. It has been noted that those subjects who ordinarily do well on Coding, very quickly make these associations.

A brief mention should be made of the eleven children in the good reading group who were also given the WISC. Inasmuch as this group consisted of so few subjects, caution should be used in drawing conclusions. It should not be inferred, for example, that there is a clear-cut picture emerging from the test results. It is interesting to note, however, that the Picture Arrangement subtest (which ranked the highest for the poor readers) ranked the lowest for this good group. There is a hint here that the good readers, in that they perhaps are more able to use abstract symbolization, may not commonly depend upon the concrete approach in learning and so are less adept in tasks where they cannot employ their peculiar abilities. The only significantly deviant subtest was Similarities which is considered to be a measure of verbal concept formation. It would be expected that good readers should do better on this type of problem. What is surprising is that the poor readers, both in this study and in Graham's, did relatively well also.

All in all, it is felt that the good readers have many strong abilities, all of them probably contributing to the learning process when applied to reading. It should be remembered that they, as a group, have in actuality no really weak areas, since their lowest subtest was only four months below the mean mental age equivalent score for the total test.

SUMMARY

In the present study, 31 poor readers and 11 good readers were given the WISC in an effort to determine whether a pattern of abilities existed which might explain the reading achievement of

each of these two groups. The achievement level was determined through the use of the reading section of the Wide-Range Achievement Test. Briefly, the findings were as follows:

1) The poor readers were significantly low on the Information, Arithmetic, and Coding subtests.

2) The poor readers were significantly high on the Picture Arrangement, Block Designs, and Comprehension subtests.

3) The good readers were significantly high on the Similarities subtest.

A hypothesis was made that the poor readers, as a group, approach learning situations in a more concrete manner as a result of an inability to handle abstractions. Since the reading process inherently consists of abstractions strongly depending on memory functions, these children are handicapped. The good readers, on the other hand, do not show this lack of ability to use abstractions and have much more retentive ability.

REFERENCES

1) Kurt Goldstein and Martin Scheerer, "Abstract and Concrete Behavior—An Experimental Study with Special Tests," *Psychological Monographs*. Vol. 53, No. 2, (1941), 151.

2) E. Ellis Graham, "Wechsler-Bellevue and Wechsler Intelligence Scale for Children Scattergrams of Unsuccessful Readers," *Journal of Consulting Psychology*. Vol. 16, (1952), 268–71.

3) J. P. Guilford, *Fundamental Statistics in Psychology and Education*. New York, McGraw-Hill Book Co., Inc., 1942.

4) Martin Mayman, Roy Schafer and David Rapaport, "Interpretation of the Wechsler Intelligence Scale in Personality Appraisal," *An Introduction to Projective Techniques*. Edited by Harold and Gladys Anderson. New York, Prentice-Hall, Inc., 1951, pp. 541–79.

5) Helen M. Robinson, *Why Pupils Fail in Reading*. Chicago, University of Chicago Press, 1946.

6) David Wechsler, *Wechsler Intelligence Scale for Children*. New York, The Psychological Corporation, 1949.

PUNITIVENESS AND "CAUSALITY" IN GRADE SCHOOL CHILDREN[1]

EUGENE E. LEVITT

Child Welfare Research Station, State University of Iowa

The preventive approach to mental health problems has been the subject of increased attention in recent years (*8*). The Preventive Psychiatry Project at the University of Iowa is one of several units currently functioning in the preventive area. The purposes and focal concepts of the project have been stated elsewhere (*1*). The present paper is a report of the investigation of one of the project's primary working hypotheses. This hypothesis may be stated thus: *In children of elementary school age, an awareness of the dynamic, complex, variable nature of human motivation is negatively related to rigid, moralistic punitiveness.* In children, this awareness does not imply a knowledge of specific motivations of human behavior, as it would in adults. It is likely that knowledge and awareness are associated even in children, although our earlier investigations have shown that there is hardly enough of the former to measure. We are assuming that it is possible for an individual who is not fully developed cognitively to realize the multiformity of causation and yet not be capable of enumerating the causal factors themselves. This realization together with its concomitants and behavioral consequences has been called "causality" (*1*). The operational definition of causality varies with cognitive development, that is, from age level to age level, although the central theme remains constant. The specific definition for the present study is based on the general formulation as it appears in the hypothesis above. It will be described in the next section.

THE TESTS

An adequate measure of punitiveness in the child, the Problem Situations Test (PST), had been developed earlier by the project. The PST is a 22-item multiple-choice test in which the child as-

---

[1] The preparation of this paper was supported by Research Grant MH-301 from the National Institute of Mental Health of the National Institutes of Health, U. S. Public Health Service. The author is indebted to Drs. William H. Lyle and Ralph H. Ojemann and to Mrs. Maxine Whiteside for assistance with various phases of the study.

sumes the hypothetical rôle of an authority figure and deals with misbehaviors or deficiencies of children. The subject is required to select one of six alternative methods of which three are punitive and three non-punitive.[2] A punitive method is one which involves physical or verbal punishment, coercion, or deprivation. The score for punitiveness is the number of punitive responses chosen. Preliminary data on the PST and its relationship to authoritarianism and parental disciplinary action will be found in another publication (*3*).

A measure of causality, the Causal Test (CT) was devised to test the hypothesis under investigation. The CT is a 30-item true-false test based on eight situations involving children. For example:

1. There is a gang of six boys in the junior high school in Sunnyside. These boys usually go around together. They are often in trouble. Once it was for tearing down a yard fence. Another time they broke all the basement windows in one of the store buildings. Recently they entered a vacant house and broke the light bulbs and tore down the window shades. What things do you think are important in explaining the behavior of these boys?

True False a. Since these boys do the same things, they are probably all alike in most ways.
True False b. A person wouldn't have to find out much about these boys to know why they do such things.
True False c. I am sure there can be no excuse for the behavior of these boys.
True False d. These boys are breaking things because they are just mean.

The test is scored inversely, i.e., for non-causality, so that the score is the number of non-causal responses to the thirty items. In twenty-four of the items, the causal response is "false". The remaining six items are distributed randomly throughout the test to prevent the establishment of a "set" for using the "false" response.

The tests were administered to 60 fourth-grade pupils, 67 fifth-grade pupils and 151 sixth graders in a town in eastern Iowa of about 80,000 population. Of the 278 subjects, 144 were boys and 134 were girls.

RESULTS

The respective mean test scores and sigmas by grade are shown in Table 1.

---

[2] These alternatives were obtained from actual responses to an earlier open-end form of the test. For further details, see (*3*).

TABLE 1—MEANS AND SD'S FOR THE TESTS

	4th Grade		5th Grade		6th Grade	
	Mean	SD	Mean	SD	Mean	SD
PST	6.91	4.48	4.38	2.72	5.39	3.76
CT	14.69	3.97	10.85	4.59	10.72	4.33

An analysis of the variance shows that there are significant grade differences on both tests. The F-ratios are 6.87 for the PST and 21.51 for the CT, both significant at well below the .01 level. On the PST, all of the individual t-test comparisons are significant. The 4th grade differs significantly from the 5th and 6th on the CT, but the latter two do not differ. Despite the differences, no clear-cut grade level trend appears. Some sort of trend seems to characterize the CT means, but no generalization concerning trend is warranted by the PST data.

In view of the significant differences between grade levels, the correlations between the two tests were computed separately for each grade. IQ scores on the Otis Self-Administering Test, Intermediate form, were secured from school records. Mean IQ's were 109 for the 4th grade, 108 for the 5th, and 105 for the 6th. Correlations between the two experimental measures and the Otis scores were also computed. All the correlations, including the relationship between PST and CT with IQ held constant, will be found in Table 2.

The correlations were averaged over the grades by means of Fisher's *z*. This procedure is justifiable, despite the independence of the groups, on the grounds that the populations are equally correlated. The correlations do not differ significantly from grade to

TABLE 2—INTERCORRELATIONS AMONG PST, CT AND IQ

Grade	PST-CT	PST-IQ	CT-IQ	PST-CT. (IQ)
4th	.41†	−.28*	−.24	.37†
5th	.52†	−.38†	−.13	.51†
6th	.45†	−.17*	−.44†	.42†
Average	.46†	−.24†	−.33†	.43†

* Significant at the .05 level.

† Significant at the .01 level.

grade except for the CT-IQ relationships for the 5th and 6th grades. Since the assumption of equally correlated populations does not hold true for CT-IQ, we shall have to regard the average CT-IQ coefficient as an expedient approximation.

DISCUSSION

Causality and punitiveness, as they have been defined in this investigation, have been shown to be significantly related in the subjects studied. The relationship is moderately high, roughly the same as has been found to exist between punitiveness and authoritarianism in children of this age group (*3*). The correlations between the two variables and intelligence are negative and significant, although relatively low. Conclusions concerning the relationship between causality and intelligence must be tentative since the average correlation is a poor estimate. Also there is no apparent explanation of the significant difference between the correlations for the 5th and 6th grades. However, we can be reasonably sure that whatever the circumstances may be, the CT-IQ relationship is not of such magnitude as to account for any large part of the correlation between causality and punitiveness. This is, of course, also true for the PST-IQ correlation.

The relationship between causality and punitiveness in children has potential importance for preventive psychiatry. We may safely assume that a child who tends to be judgmentally punitive, whether this tendency is overt or suppressed, is inclining in an unwholesome direction from the point of view of mental hygiene. Various studies in the area provide evidence that this type of aggressiveness may be an etiological factor in race prejudice, delinquency, and character disorders. While our data suggest a general relationship between punitiveness and causality they do not indicate which of the variables is the prime mover and which is the effect. The purposes of prevention do not require that the matter actually be clarified. The pertinent question asks whether punitiveness in the child can be reduced by means of school curricula and other planned learning programs designed to bring about an increase in causality. Certain investigations suggest that this is the case. Stiles (*7*) has shown that even a brief learning program resulted in a decrease in punitiveness in 10-year olds as measured by their treatment of peer problems in the "room council." Links (*2*) found that an "analytic" (causal) radio program was more effective than a non-analytic one in reducing anti-Negro prejudice in 10th grade pupils. Indirect evidence is furnished by supplementary studies (*4*, *6*).

A general outline of the learning program for teaching causality as developed thus far has been presented by Ojemann (*5*). Manuals and teaching aids developed within this framework are available to educators and other interested persons on request.

REFERENCES

1) E. E. Levitt and R. H. Ojemann, "The aims of preventive psychiatry and 'causality' as a personality pattern." *J. Psychol.*, 1953, 36: 393–400.

2) Eleanor Links, *A comparison of two methods of using radio in the changing of certain social attitudes*. Unpublished Master's thesis, State Univ. of Iowa, 1945.

3) W. H. Lyle and E. E. Levitt, "Punitiveness, authoritarianism and parental discipline of grade school children." *J. abnorm. soc. Psychol.*, 1955, 51: 42–46.

4) Mildred I. Morgan and R. H. Ojemann, "The effect of a learning program designed to assist youth in an understanding of behavior and its development." *Child Develpmt.*, 1942, 13: 181–194.

5) R. H. Ojemann, "An integrated plan for education in human relations and mental health." *J. schl. Hlth.*, 1950, 20: 99–106.

6) R. H. Ojemann and Frances R. Wilkinson, "The effect on pupil growth of an increase in teacher's understanding of pupil behavior." *J. exp. Educ.*, 1939, 9: 143–147.

7) Frances S. Stiles, "Developing an understanding of human behavior at the elementary school level." *J. educ. Res.*, 1950, 43: 516–524.

8) *Promotion of mental health in the primary and secondary schools: an evaluation of four projects*. Report No. 18, Committee on Preventive Psychiatry of the Group for the Advancement of Psychiatry, January, 1951.

A FORCED CHOICE RATING FORM FOR COLLEGE TEACHERS

ROLLIN M. PATTON and PRISCILLA R. MEYER

The University of Akron

Over 225 colleges and universities are now using student ratings of instructors and, at last count, another 284 were either considering it or had definite plans made (*4*). Faculty resistance to such ratings is well known. The primary objections are: (a) proneness of the ratings to leniency and halo; (b) carelessness and superficiality of student ratings; (c) use of generalized traits for rating; and (d) disruption of normal student-teacher relations.

Faculty members often think that these difficulties are peculiar to classroom ratings. They may not realize that these effects appear in many rating situations, and can be minimized by the use of better scales.

The purpose of this study was to produce a scale which would answer some or all of these objections. The forced choice method was selected because it minimizes halo and leniency, yields ratings in terms of specific behavior, requires care and discrimination by the rater, and does not require the use of unfavorable statements.

THE PRELIMINARY SCALE

A list of 244 statements describing specific teacher behavior was compiled from essays about "My Best College Teacher," and "My Worst College Teacher" written by 50 college students. All statements were re-written to appear favorable. For example, "He puts the class to sleep" was changed to "He keeps the class awake."

Eighty-eight students rated each statement on a five-point scale indicating degree of applicability to a teacher whom they had known in the past. Half were asked to rate the best teacher whom they had ever known, half the worst teacher. From the ratings a discrimination index for each statement was computed according to the method of Sisson (*7*).

Nine students not included in the first group rated the statements on a five-point scale of favorability of appearance. Their ratings agreed closely, and the mean rating was used as the favorability index. This method, suggested by Berkshire and Highland

(*1*), was employed in place of the more usual preference index. Only items with a mean rating of moderately- to highly-favorable were retained.

Thirty-one forced choice blocks were constructed by placing in each block four items with the same favorability index, two of which had high discrimination indices and two of which discriminated poorly. Subjects were instructed to select two items from each block that best described the teacher to be rated.

At University A the scale was administered to 55 students who rated the best teacher that they had ever had in college, and 54 who rated the worst. The sample was taken from 10 classes in engineering, liberal arts, business and education. Ratings were anonymous but information collected on the form showed that over half the subjects had attended two or more colleges. No freshmen were used.

REFINING THE PRELIMINARY SCALE

After item analysis the six poorest blocks were dropped. The remaining 25 blocks, containing 100 items, covered the following areas: method of teaching, examination and grading procedures, knowledge of the subject, method of handling students' questions, student-teacher relations, textbooks and assignments, speech habits and personal traits. The items were scored as follows: +2 for items discriminating good teachers; +1 for non-discriminating items; 0 for items discriminating poor teachers. Possible scores ranged from 20 to 84. Administration required about 15 minutes.

TRYOUT AND CROSS VALIDATION

The revised scale, called the *Instructor's Information Form*,[1] was administered at two additional schools. At University B, a private urban school with 10,000 students, it was given to 98 students from seven classes. At University C, a rural state school with 5000 students, it was given to 110 students from four classes. Of the total sample of 208, 28% were sophomores, 29% juniors, 31% seniors and 12% graduate students. Half the sample rated their best teacher, half their worst, using teachers from any college they had attended in the last four years. Because the data were

---

[1] The *Instructor's Information Form* may be obtained from Dr. Rollin M. Patton, 535 Blakeley Drive, Antonio 9, Texas.

TABLE I—SUMMARY OF RESULTS

	N	Best Teachers		Worst Teachers		t
		Mean	S.D.	Mean	S.D.	
University B	98	66.76	4.28	43.76	7.04	19.00*
University C	110	65.02	4.77	46.26	6.48	17.53*
Combined Groups	208	65.92	4.62	44.22	6.99	26.20*

* Significant at .01 level of confidence.

obtained in summer sessions an unusually cosmopolitan student sample resulted. Over 57 % had attended other colleges; they reported attendance at 73 different schools well scattered across the country.

Table I shows the means of the ratings. All differences between means of best and worst teachers are significant at the .01 level of confidence (t-test). Differences between the means of best teachers at the two universities are not significant, nor are differences between worst teachers. The frequency distributions for the best and worst groups are shown in Table II, with the data for Universities B and C combined.

The split-half reliability of the scale, calculated on odd-even items, was found to be 0.88.

RESISTANCE TO LENIENCY AND HALO

Success in biasing ratings requires the ability to identify the high and low scoring items. Such identification was very easy on three teacher rating scales examined by the authors (*2*, *5*, *6*). On two of these, graphic scales, the highest rating was always on the right; on the other, a check list, the good items were always stated in favorable terms, the poor ones in unfavorable terms. Spontaneous comments by our subjects have indicated that good and poor items on the *Instructor's Information Form* are not readily discriminated.

For an estimate of the scale's resistance to faking, 39 students from the authors' classes in psychology were asked to fill it out with the highest possible rating. If the scale could be falsified perfectly, the mean of the faking group should be 84 (the highest possible score). Actually the mean of the faking group was 63.90 and the standard deviation 5.32. The highest faked rating was 72.

TABLE II—FREQUENCY DISTRIBUTIONS OF THE SCORES OBTAINED WHEN 104 STUDENTS RATED THEIR BEST TEACHER AND 104 STUDENTS RATED THEIR WORST TEACHER (DATA FOR UNIVERSITIES B AND C COMBINED)

Score	Worst Teachers	Best Teachers
31–32	2	
33–34	5	
35–36	5	
37–38	8	
39–40	6	
41–42	6	
43–44	13	
45–46	13	
47–48	8	
49–50	9	
51–52	13	
53–54	9	
55–56	4	1
57–58	3	4
59–60		10
61–62		9
63–64		13
65–66		21
67–68		19
69–70		11
71–72		7
73–74		6
75–76		1
77–78		1
79–80		1

The faked mean is 2.02 below the mean for best teachers, a small difference which was, however, significant at the .05 level of confidence (t-test).

CONCLUSIONS

A scale of the forced choice type shows ability to discriminate good and poor teachers with very little overlapping. It is believed that such a scale, consisting of favorable items describing specific teacher behavior, can answer faculty complaints that student ratings are too general, too carelessly made, or too embarrassing. The forced choice scale shows greater resistance to faking than other types.

REFERENCES

1) J. R. Berkshire and R. W. Highland, "Forced Choice Rating." *Personnel Psychology*, 1953, 6: 355–378.

2) C. W. Crannell, "Experiment in Rating of Instructors by Students." *College and University*, 1948, 24: 5–11.

3) P. R. Meyer and R. M. Patton, "Can Student Rating of Instructors be Painless and Foolproof?" *School and Society*, 1954, 80: 200–201.

4) F. J. Mueller, "Trends in Student Ratings of Faculty." *Bulletin of the American Association of University Professors*, 1951, 37: 319–324.

5) H. H. Remmers and P. C Baker, *The Purdue Rating Scale for Instruction.* Lafayette, Indiana; Division of Educational Reference, Purdue University, 1950.

6) H. Ruja, "Instructor-Rating Scale." *Educational Administration and Supervision*, 1953, 39: 209–217.

7) E. D. Sisson, "New Army Rating." *Personnel Psychology*, 1948, 1: 365–381.